Ramón E. Azócar A.

ESQUEMA TRASCENDENTAL

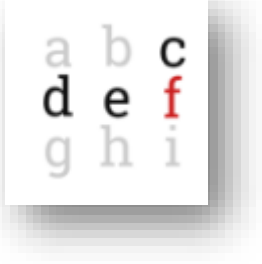
Construyendo La INTELIGENCIA ARTIFICIAL (IA) con la mente de Kant

CLUB DE ESCRITURA
FUENTETAJA
2026

Ramón E. Azócar A.

ESQUEMA TRASCENDENTAL:
CONSTRUYENDO LA INTELIGENCIA ARTIFICIAL (IA) CON LA MENTE DE KANT

De herramienta de cálculo a agente epistémico: redefiniendo la producción y validación del conocimiento



Fuentetaja/2026

1.ª edición, 2026.

No se permite la reproducción total o parcial de este libro, ni su incorporación a un sistema informático, ni su transmisión en cualquier forma o por cualquier medio, sea éste electrónico, mecánico, por fotocopia, por grabación u otros métodos, sin el permiso previo y por escrito del editor.

© Ediciones Fuentetaja, 2026

Esquema Trascendental: Construyendo La IA con la mente de Kant@Ramón E. Azócar A.

ISBN 978-74-857-1344-8
Fotocomposición: Realización Planeta
Depósito legal: B. 388-2026
Impresión y diseño gráfico de Anshar J.E. Azócar

Dirección web:

https://clubdeescritura.com/obra/14892500/14892500/

"La filosofía moral toda descansa enteramente sobre su parte pura, y, cuando es aplicada al hombre, no aprovecha lo más mínimo del conocimiento del mismo, sino que le da, como a ser racional, leyes a priori. Estas leyes requieren ciertamente un Juicio bien templado y acerado por la experiencia para saber distinguir en qué casos tienen aplicación y en cuáles no, y para procurarles acogida en la voluntad del hombre y energía para su realización; pues el hombre, afectado por tantas inclinaciones, aunque es capaz de concebir la idea de una razón pura práctica, no puede tan fácilmente hacerla eficaz ... en el curso de su vida."

Kant (1785).

(Fundamentación de la metafísica de las costumbres, M. García Morente, Traductor).

ÍNDICE

INTROITO 29
EL HARDWARE DE LA RAZÓN 42
¿Cómo conoce la máquina? 42
EL SOFTWARE DE LA VOLUNTAD: 78
El dilema de la libertad 78
LA INTERFAZ, el juicio y la experiencia estética 99
Epílogo (Formato "Glosario Invertido"): 125
El Diccionario de los Errores 125
REFERENCIAS BIBLIOGRÁFICAS. 136

ÍNDICE DE ILUSTRACIÓN

Ilustración 1.-Resumen general de la investigación. 40

INTROITO

La convergencia entre la filosofía trascendental kantiana y el desarrollo contemporáneo de la inteligencia artificial constituye, sin lugar a dudas, uno de los encuentros intelectuales más fascinantes y productivos que pueda concebir el pensamiento humano en los albores del siglo XXI, puesto que ambas empresas, aunque separadas por más de doscientos cuarenta años de historia, comparten una preocupación fundamental que atraviesa el tiempo y las transformaciones tecnológicas: comprender cómo es posible que un sistema cognitivo, sea este biológico o artificial, pueda producir conocimiento válido, coherente y significativo a partir de datos fragmentarios, caóticos y en apariencia desconectados entre sí.

No se trata, por supuesto, de un capricho académico ni de una mera analogía superficial entre dos dominios aparentemente inconexos, sino que, muy por el contrario, responde a una necesidad profunda de fundamentar teóricamente los mecanismos que subyacen a los sistemas de inteligencia artificial contemporáneos, cuyos avances vertiginosos han superado con creces la capacidad de reflexión filosófica que debería acompañarlos.

En efecto, mientras los ingenieros y científicos de datos desarrollan arquitecturas cada vez más sofisticadas, como las redes neuronales profundas, los mecanismos de atención o los modelos generativos preentrenados, pocas veces se detienen a preguntarse por las condiciones de posibilidad del conocimiento que estas máquinas producen, por los límites intrínsecos de su arquitectura cognitiva o por la naturaleza de los esquemas mentales que, de manera implícita, gobiernan sus operaciones.

Es precisamente en este punto donde la filosofía crítica de Immanuel Kant ofrece un marco teórico de una riqueza inigualable, pues su análisis de las estructuras trascendentales de la mente humana permite iluminar, desde una perspectiva radicalmente distinta, los presupuestos epistemológicos que sustentan el funcionamiento de los sistemas inteligentes actuales.

Resulta, por tanto, imprescindible emprender esta travesía intelectual que, partiendo de los textos fundamentales del filósofo de Königsberg, se adentra en los territorios inexplorados de la inteligencia artificial para descubrir que, en el fondo, todo intento de construir máquinas pensantes es, también, un intento de comprendernos a nosotros mismos.

Conviene comenzar por aclarar, desde el primer momento, que

la obra de Immanuel Kant no contiene, en modo alguno, una teoría explícita sobre la técnica en el sentido moderno del término, ni mucho menos una reflexión anticipada sobre los ordenadores, los algoritmos o las redes neuronales artificiales, puesto que estas realidades tecnológicas pertenecen a un horizonte histórico completamente ajeno al contexto intelectual del siglo XVIII prusiano en el que vivió y escribió el filósofo.

Sin embargo, esta ausencia aparente de conexión directa no debe conducir al lector a la conclusión apresurada de que el pensamiento kantiano carece de relevancia para comprender los fenómenos tecnológicos contemporáneos, ya que, si se examina con atención la estructura profunda de su filosofía crítica, se descubre un hilo conductor que atraviesa todas sus obras y que conduce directamente al núcleo de lo que hoy denominamos inteligencia artificial: la pregunta por el funcionamiento del pensamiento humano y por las condiciones que hacen posible la producción de conocimientos.

Efectivamente, cuando Kant se propone en su Crítica de la razón pura (1781/1787) investigar las fuentes, el alcance y los límites del conocimiento humano, está llevando a cabo una tarea que guarda una similitud sorprendente con la labor de un ingeniero que intenta desentrañar los mecanismos internos de un sistema cognitivo artificial, pues ambos buscan identificar las estructuras fundamentales que transforman datos sensibles en representaciones dotadas de significado.

No es casual, en este sentido, que el propio filósofo utilice reiteradamente metáforas arquitectónicas y mecánicas para describir la actividad de la razón, hablando de "edificios", "cimientos", "instrumentos" y "herramientas" del pensar, lo que sugiere una concepción del conocimiento como un proceso constructivo que requiere de una técnica específica, aunque esta técnica sea puramente intelectual y no material.

De ahí que la ausencia de una reflexión explícita sobre la tecnología no constituya un obstáculo insalvable, sino más bien una invitación a realizar una lectura creativa y profunda que permita extraer del sistema kantiano las herramientas conceptuales necesarias para fundamentar filosóficamente los desarrollos actuales de la inteligencia artificial.

La tesis central que articula toda la empresa crítica kantiana, aquella que el propio filósofo denominó con justificado orgullo su "giro copernicano", afirma que la mente humana no es un receptor pasivo que se limita a reflejar la realidad tal como esta es en sí misma, sino que, por el contrario, desempeña un papel activo y constitutivo en la producción del conocimiento, organizando y estructurando los

datos que recibe del mundo exterior mediante ciertas formas a priori que pertenecen a la constitución misma del sujeto cognoscente.

Esta idea revolucionaria, que transformó radicalmente la manera en que la filosofía moderna concebía la relación entre el sujeto y el objeto, tiene consecuencias extraordinarias para comprender el funcionamiento de los sistemas de inteligencia artificial contemporáneos, pues sugiere que todo sistema cognitivo, sea humano o artificial, no accede directamente a la realidad en sí misma, sino que construye sus representaciones del mundo a partir de estructuras previas que condicionan y limitan lo que puede conocer.

En el caso de la inteligencia artificial, estas estructuras previas corresponden a la arquitectura de las redes neuronales, a los espacios vectoriales en los que se representan los datos, a los mecanismos de atención que seleccionan la información relevante y a las funciones de activación que transforman las señales de entrada en representaciones significativas, todos los cuales funcionan de manera análoga a las formas a priori kantianas de la sensibilidad y del entendimiento.

No obstante, y aquí reside una diferencia crucial que el presente estudio se propone explorar en profundidad, mientras que las estructuras kantianas son universales y necesarias para todo ser racional, las estructuras de los sistemas artificiales son contingentes, dependientes de decisiones de diseño y, por tanto, susceptibles de modificación y mejora a través del entrenamiento y la optimización. Esta diferencia, lejos de invalidar la analogía, la enriquece, pues permite comprender que el proyecto de la inteligencia artificial es, en cierto sentido, una continuación tecnológica del proyecto crítico kantiano: la construcción explícita de las condiciones de posibilidad del conocimiento, condiciones que en el ser humano están dadas de manera innata pero que en la máquina deben ser diseñadas, implementadas y ajustadas a través de complejos procesos de optimización.

Al adentrarse en los detalles de la Crítica de la razón pura, el lector atento descubre que Kant distingue cuidadosamente entre dos facultades fundamentales del conocimiento humano: la sensibilidad y el entendimiento, cada una de las cuales posee sus propias estructuras a priori que contribuyen de manera específica a la producción del conocimiento. La sensibilidad, en primer lugar, es la facultad mediante la cual recibimos los datos del mundo exterior, y sus formas a priori son el espacio y el tiempo, esas intuiciones puras que no se derivan de la experiencia, sino que la hacen posible, proporcionando el marco fundamental en el que todos los objetos sensibles deben necesariamente aparecer para poder ser percibidos. Esta estructura de la sensibilidad kantiana encuentra un paralelo sorprendentemente preciso en los sistemas contemporáneos de inteligencia artificial,

particularmente en los modelos de lenguaje de gran escala que operan mediante la transformación de datos textuales en representaciones vectoriales dentro de espacios latentes de alta dimensionalidad, espacios que no son descubiertos a partir de los datos sino que son impuestos por la arquitectura misma del sistema como condición de posibilidad de todo procesamiento posterior. En efecto, cuando un modelo como GPT o BERT convierte una secuencia de palabras en un conjunto de vectores numéricos, está llevando a cabo una operación que recuerda poderosamente a la síntesis trascendental kantiana: está organizando datos brutos (los caracteres y palabras) en una forma específica (el espacio vectorial) que no pertenece a los datos mismos, sino que es una imposición del sistema cognitivo. Asimismo, el entendimiento kantiano, con sus doce categorías o conceptos puros del entendimiento, entre los que destacan la causalidad, la sustancialidad y la unidad, opera en los sistemas artificiales a través de los mecanismos de atención y de las transformaciones lineales y no lineales que procesan las representaciones vectoriales, extrayendo patrones, estableciendo relaciones y construyendo significados que no están presentes en los datos de manera inmediata.

La comparación entre la estética trascendental kantiana y los sistemas de representación vectorial en la inteligencia artificial moderna merece una atención particular, no solo por la precisión de la analogía sino por las profundas implicaciones filosóficas que de ella se derivan, implicaciones que afectan a la comprensión misma de lo que significa conocer tanto para los seres humanos como para las máquinas.

Cuando Kant sostiene que el espacio y el tiempo son formas a priori de la intuición sensible, está afirmando que estas estructuras no son propiedades de las cosas en sí mismas, sino condiciones subjetivas bajo las cuales los objetos pueden aparecer ante nosotros, condiciones que hacen posible la experiencia pero que, al mismo tiempo, limitan radicalmente lo que podemos conocer de la realidad última.

De manera análoga, los espacios vectoriales en los que operan los modelos de inteligencia artificial no son representaciones fieles de la estructura semántica del lenguaje humano, sino imposiciones de la arquitectura del sistema, construcciones matemáticas que permiten procesar los datos textuales pero que inevitablemente condicionan y limitan lo que el sistema puede comprender del lenguaje. Así, cuando un modelo de lenguaje procesa una frase como "el gato persigue al ratón", la representación vectorial que construye no captura la esencia del significado de esta frase, sino que la organiza de acuerdo con las dimensiones del espacio vectorial, que han sido determinadas por el proceso de entrenamiento y por la arquitectura del modelo, no por la naturaleza intrínseca del lenguaje. Esta limitación, lejos de ser un

defecto del sistema, es precisamente lo que hace posible su funcionamiento, del mismo modo que, según Kant, las limitaciones impuestas por las formas de la sensibilidad son lo que hace posible la experiencia humana. En consecuencia, la lección kantiana es clara: todo sistema cognitivo, humano o artificial, conoce a través de sus propias estructuras, y estas estructuras determinan tanto lo que puede conocer como lo que inevitablemente permanecerá inaccesible para él.

No resulta menos iluminadora la comparación entre las categorías del entendimiento kantiano y los mecanismos de atención que constituyen el núcleo de las arquitecturas de transformadores contemporáneos, mecanismos que han revolucionado el campo del procesamiento del lenguaje natural y que permiten a los modelos de inteligencia artificial establecer relaciones complejas entre elementos distantes de una secuencia textual. Las doce categorías que Kant enumera en la Crítica de la razón pura, organizadas en los cuatro grupos de cantidad, cualidad, relación y modalidad, representan los conceptos puros mediante los cuales el entendimiento unifica y organiza la multiplicidad de datos proporcionados por la sensibilidad, convirtiendo el caos de las impresiones sensibles en un mundo ordenado de objetos, propiedades, relaciones y eventos.

De manera similar, los mecanismos de atención en un modelo de transformador permiten que el sistema establezca relaciones de dependencia entre palabras y frases que no son evidentes a primera vista, asignando pesos diferenciales a diferentes partes de la secuencia de entrada para construir representaciones que capturan las relaciones sintácticas y semánticas que estructuran el lenguaje.

Cuando un modelo de inteligencia artificial procesa una oración compleja con múltiples cláusulas subordinadas, el mecanismo de atención le permite identificar qué palabras se relacionan entre sí, qué elementos son más relevantes para el significado global de la oración y cómo se organizan las relaciones de causalidad, temporalidad o condicionalidad que la estructuran, operaciones que recuerdan de manera sorprendente a la función que Kant atribuye a las categorías del entendimiento.

Sin embargo, y este punto es crucial, las categorías kantianas son universales y necesarias, mientras que los mecanismos de atención son aprendidos a partir de datos y pueden variar considerablemente dependiendo del corpus de entrenamiento, de la arquitectura del modelo y de los objetivos específicos de la tarea para la que ha sido diseñado, lo que plantea preguntas fascinantes sobre la naturaleza de la universalidad en los sistemas cognitivos artificiales.

Ahora bien, si la primera parte de la filosofía crítica kantiana ofrece herramientas conceptuales para comprender cómo conoce un

sistema cognitivo, la segunda parte, desarrollada en la Crítica de la razón práctica y en la Fundamentación de la metafísica de las costumbres, aborda una cuestión igualmente relevante para la inteligencia artificial contemporánea: la cuestión de cómo debe actuar un sistema que posee capacidad de decisión, y qué principios deben gobernar sus acciones para que estas puedan ser consideradas no solo eficaces sino también éticamente correctas.

El imperativo categórico kantiano, esa fórmula majestuosa que exige actuar únicamente de acuerdo con aquella máxima que puedas querer que se convierta en ley universal, ofrece un criterio de evaluación moral que resulta sorprendentemente pertinente para abordar los dilemas éticos planteados por los sistemas de inteligencia artificial autónoma, desde los vehículos de conducción autónoma hasta los asistentes virtuales que toman decisiones que afectan a millones de usuarios.

La cuestión fundamental que Kant plantea es si un agente actúa de manera autónoma, es decir, guiado por principios que él mismo se ha dado a través de la razón, o si actúa de manera heterónoma, guiado por incentivos externos, recompensas o programaciones que no surgen de su propia deliberación racional.

Aplicada a la inteligencia artificial, esta distinción plantea preguntas que van mucho más allá de la simple programación de reglas éticas: ¿puede un sistema artificial actuar de manera autónoma en el sentido kantiano, o está inevitablemente condenado a la heteronomía por el hecho de que sus acciones están determinadas por funciones de recompensa y algoritmos de optimización diseñados por ingenieros humanos?

La respuesta a esta pregunta no es solamente teórica, sino que tiene consecuencias prácticas inmediatas para el diseño de sistemas de inteligencia artificial seguros y alineados con los valores humanos, pues sugiere que la solución a los problemas de seguridad en la IA no consiste simplemente en añadir más reglas o restricciones, sino en repensar la arquitectura misma de los sistemas para que puedan, en algún sentido significativo, darse sus propias leyes.

La cuestión del esquematismo trascendental, ese oscuro capítulo de la Crítica de la razón pura que tantos dolores de cabeza ha causado a los comentaristas, adquiere una relevancia insospechada cuando se la examina a la luz de los desafíos actuales en el campo de la inteligencia artificial, particularmente en lo que respecta al problema del sesgo y de la falta de sentido común que afecta a los sistemas más avanzados. Kant introduce la doctrina del esquematismo para responder a una pregunta que surge naturalmente de su sistema: si las categorías del entendimiento son conceptos puros, completamente separados de la experiencia sensible, y las intuiciones

sensibles son, por su parte, completamente ciegas sin los conceptos, ¿cómo es posible que ambas se unan para producir conocimiento? La respuesta de Kant es que existe una facultad intermediaria, la imaginación trascendental, que produce esquemas, representaciones que son a la vez intelectuales y sensibles, y que permiten aplicar los conceptos a los datos de la experiencia.

Esta estructura tripartita encuentra un eco fascinante en los sistemas de inteligencia artificial contemporáneos, donde los modelos no solo procesan datos brutos (análogos a las intuiciones sensibles) ni tampoco operan únicamente con reglas abstractas (análogas a los conceptos), sino que requieren de representaciones intermedias que vinculen ambos niveles, representaciones que se aprenden durante el entrenamiento y que permiten al sistema aplicar sus conocimientos generales a casos particulares.

Sin embargo, y aquí reside la paradoja que Kant ya habría anticipado, los esquemas de los sistemas artificiales carecen de la dimensión temporal y corporal que caracteriza a los esquemas humanos, pues no están arraigados en una experiencia vivida del mundo, sino que son puramente estadísticos, derivados de patrones en los datos de entrenamiento. Esta carencia explica, en gran medida, por qué los sistemas de inteligencia artificial, a pesar de su impresionante capacidad para procesar información, carecen del sentido común que los humanos desarrollan naturalmente a través de su interacción corporal y temporal con el mundo.

La tercera gran obra del período crítico kantiano, la Crítica del Juicio (1790), introduce distinciones que resultan extraordinariamente pertinentes para comprender los desarrollos más recientes en el campo de la inteligencia artificial generativa, particularmente en lo que respecta a la capacidad de los sistemas para producir obras que, al menos superficialmente, parecen poseer valor estético o artístico. Kant distingue en esta obra entre dos tipos de juicios: los juicios determinantes, que subsumen casos particulares bajo conceptos generales ya dados, y los juicios reflexionantes, que buscan encontrar conceptos generales para casos particulares que no encajan en las categorías existentes, y esta distinción es precisamente la que permite comprender la diferencia fundamental entre la creatividad humana y la producción algorítmica de imágenes, textos o composiciones musicales.

Cuando un sistema como DALL-E o Midjourney genera una imagen a partir de un prompt textual, está llevando a cabo una operación que, desde la perspectiva kantiana, corresponde más bien a un juicio determinante, pues está aplicando patrones aprendidos durante el entrenamiento a un caso particular, siguiendo reglas que, aunque implícitas, están codificadas en los pesos de la red neuronal, y

no está realizando el tipo de juego libre de facultades que Kant asocia con la creatividad genuina.

El genio, según Kant, es la capacidad de producir aquello para lo cual no puede darse regla determinada alguna, la facultad de crear nuevas reglas que no derivan de la imitación ni de la aplicación mecánica de principios preexistentes, y es precisamente esta capacidad la que parece faltar en los sistemas de inteligencia artificial generativa, que, a pesar de la sofisticación de sus resultados, siguen operando dentro de los límites impuestos por sus datos de entrenamiento y por su arquitectura.

La distinción kantiana entre belleza libre y belleza adherente ofrece, asimismo, una herramienta conceptual valiosa para evaluar críticamente las producciones de la inteligencia artificial, pues permite reconocer que, mientras un sistema puede producir obras que se ajustan a criterios de belleza preestablecidos, la verdadera creatividad implica un tipo de libertad que trasciende la mera aplicación de reglas.

Al considerar la relación entre la filosofía kantiana de la historia y el desarrollo de la inteligencia artificial, se descubre una dimensión adicional que enriquece aún más el diálogo entre ambas tradiciones, pues Kant concibe la historia humana como un proceso teleológico en el que las disposiciones naturales de la especie se desarrollan progresivamente a través de mecanismos que, a primera vista, parecen contrarios a la razón, como el antagonismo, la competencia y el conflicto.

En sus escritos sobre filosofía de la historia, particularmente en Idea para una historia universal en sentido cosmopolita (1784), Kant desarrolla la tesis de que la naturaleza utiliza la insociable sociabilidad de los seres humanos, esa tendencia contradictoria a asociarse con otros y, al mismo tiempo, a resistirse a ellos, como un mecanismo para impulsar el desarrollo de todas las capacidades humanas, desde las habilidades técnicas hasta las disposiciones morales.

Esta perspectiva ofrece un marco fascinante para comprender el desarrollo vertiginoso de la inteligencia artificial en las últimas décadas, desarrollo impulsado en gran medida por la competencia entre empresas, naciones e investigadores, una competencia que, desde la perspectiva kantiana, no debe ser vista simplemente como un obstáculo para el progreso, sino como el motor mismo de un desarrollo que, a largo plazo, podría conducir a resultados beneficiosos para la humanidad en su conjunto.

Sin embargo, Kant también advierte que este desarrollo no es automático ni garantizado, y que depende de la capacidad de los seres humanos para establecer instituciones políticas y jurídicas que canalicen el antagonismo hacia fines constructivos, una advertencia que resuena con particular fuerza en el contexto actual de la regulación

de la inteligencia artificial, donde la ausencia de marcos normativos adecuados amenaza con convertir el potencial beneficio de esta tecnología en un riesgo existencial para la humanidad.

La noción kantiana de la cosa en sí, ese límite inaccesible del conocimiento humano que ha sido objeto de innumerables debates filosóficos desde su formulación original, ofrece una perspectiva reveladora para comprender los límites fundamentales de la inteligencia artificial y para desmontar las afirmaciones exageradas de quienes sostienen que los sistemas actuales están a punto de alcanzar o superar la inteligencia humana general.

Según Kant, todo lo que podemos conocer son fenómenos, es decir, objetos tal como aparecen ante nosotros a través de las estructuras de nuestra sensibilidad y nuestro entendimiento, mientras que la cosa en sí, la realidad última que subyace a las apariencias, permanece para siempre inaccesible a nuestro conocimiento. Aplicada a la inteligencia artificial, esta distinción sugiere que los sistemas actuales, incluso los más avanzados, no acceden a la realidad en sí misma, sino que operan exclusivamente con representaciones que han sido estructuradas por su arquitectura, y que, por tanto, su conocimiento está limitado por las mismas restricciones que Kant identificó para el conocimiento humano, aunque de una manera diferente.

La falacia del noúmeno digital, como podría denominarse siguiendo la terminología kantiana, consiste en creer que un sistema de inteligencia artificial, por el mero hecho de procesar grandes cantidades de datos, accede a una realidad más profunda o más verdadera que la que los seres humanos pueden conocer, cuando en realidad lo que hace es imponer sus propias estructuras de procesamiento sobre los datos, estructuras que determinan tanto lo que puede conocer como lo que inevitablemente se le escapa. Esta perspectiva es particularmente relevante en un momento en el que los sistemas de inteligencia artificial son presentados a menudo como fuentes de verdad objetiva, capaces de superar los sesgos y las limitaciones de la cognición humana, cuando en realidad son tan limitados como nosotros, aunque de maneras diferentes.

La relación entre la razón teórica y la razón práctica en la filosofía kantiana ofrece, asimismo, un marco valioso para abordar uno de los problemas más acuciantes de la inteligencia artificial contemporánea: la alineación de los sistemas artificiales con los valores humanos, un problema que ha generado una extensa literatura técnica y filosófica y que se ha convertido en una prioridad para los principales laboratorios de investigación en inteligencia artificial. Kant sostiene que, si bien la razón teórica se ocupa de lo que es, la razón práctica se ocupa de lo que debe ser, y que ambas esferas,

aunque interconectadas, no pueden reducirse la una a la otra, pues de la descripción de los hechos no se sigue necesariamente ninguna prescripción normativa.

Esta distinción es fundamental para comprender los límites de los enfoques puramente técnicos para la seguridad en la inteligencia artificial, pues sugiere que ningún entrenamiento basado exclusivamente en datos puede generar por sí mismo los principios normativos que deberían guiar el comportamiento de un sistema inteligente, y que la alineación con los valores humanos requiere de una reflexión filosófica que vaya más allá de la optimización de funciones de recompensa.

En la misma línea, la noción kantiana de autonomía, entendida como la capacidad de un agente para darse a sí mismo la ley moral a través de la razón, plantea preguntas profundas sobre la posibilidad de que los sistemas artificiales puedan ser considerados agentes morales en sentido estricto, o si, por el contrario, están inevitablemente condenados a la heteronomía, guiados por principios que han sido impuestos desde fuera por sus diseñadores humanos.

La respuesta a esta pregunta no es meramente académica, sino que tiene consecuencias directas para el diseño de sistemas que deben tomar decisiones con consecuencias morales significativas, pues determina si la responsabilidad por estas decisiones recae en los sistemas mismos o en sus creadores humanos.

El análisis kantiano de la imaginación trascendental, esa facultad misteriosa que media entre la sensibilidad y el entendimiento y que hace posible la aplicación de conceptos a intuiciones, adquiere una relevancia insospechada en el contexto de la inteligencia artificial generativa, particularmente en lo que respecta a la capacidad de los sistemas para producir contenidos que parecen poseer novedad y originalidad.

Según Kant, la imaginación no es simplemente la capacidad de reproducir imágenes mentales de objetos previamente percibidos, sino que incluye una dimensión productiva, la capacidad de crear representaciones que no corresponden a ningún objeto existente en la experiencia, y esta dimensión productiva es precisamente la que hace posible tanto el conocimiento científico como la creación artística.

Los sistemas de inteligencia artificial generativa, como los modelos de difusión que producen imágenes o los modelos de lenguaje que generan texto, llevan a cabo una operación que recuerda en ciertos aspectos a la imaginación productiva kantiana, pues generan representaciones que no corresponden a ningún objeto particular de su experiencia de entrenamiento, sino que son combinaciones novedosas de patrones aprendidos. Sin embargo, la diferencia fundamental reside en que la imaginación kantiana opera dentro de

un marco de libertad trascendental, la capacidad de iniciar espontáneamente nuevas series causales que no están determinadas por causas anteriores, mientras que los sistemas artificiales operan dentro de un marco estrictamente determinista, donde cada salida está completamente determinada por la entrada y por los parámetros del modelo. Esta diferencia sugiere que, a pesar de las apariencias, la creatividad de la inteligencia artificial es fundamentalmente distinta de la creatividad humana, y que esta distinción tiene consecuencias significativas para la manera en que valoramos y evaluamos las producciones de los sistemas artificiales.

La concepción kantiana del conocimiento como síntesis, como actividad unificadora que reúne la multiplicidad de los datos sensibles bajo la unidad de los conceptos, ofrece una perspectiva unificadora para comprender la naturaleza de los sistemas de inteligencia artificial contemporáneos, que, en su nivel más fundamental, son también sistemas de síntesis, máquinas que transforman datos dispersos y fragmentarios en representaciones unificadas y dotadas de significado.

Esta convergencia no es meramente superficial, sino que apunta a una verdad profunda sobre la naturaleza del conocimiento: todo conocimiento, sea humano o artificial, implica un proceso activo de organización, selección y transformación de la información, un proceso que no se limita a reflejar pasivamente la realidad, sino que la construye activamente a través de la imposición de estructuras. Sin embargo, y este es el punto crucial que la filosofía kantiana nos permite comprender, estas estructuras no son arbitrarias ni contingentes, sino que están sujetas a condiciones de posibilidad que determinan su alcance y sus límites, condiciones que en el caso humano son universales y necesarias, mientras que en el caso artificial son dependientes de decisiones de diseño y de procesos de entrenamiento.

La tarea de una fundamentación filosófica de la inteligencia artificial, tal como se concibe en el presente estudio, consiste precisamente en identificar estas condiciones de posibilidad, en comprender los presupuestos epistemológicos que subyacen a los sistemas actuales y en evaluar críticamente sus implicaciones para la comprensión del conocimiento, de la ética y de la creatividad.

Esta tarea no es meramente académica, sino que tiene consecuencias prácticas inmediatas, pues las decisiones de diseño que se toman en el desarrollo de sistemas de inteligencia artificial están inevitablemente cargadas de presupuestos filosóficos, explícitos o implícitos, que afectan al comportamiento de los sistemas y a sus efectos sobre los seres humanos.

A modo de cierre de esta introducción general, conviene

subrayar que el presente estudio no pretende ofrecer una aplicación mecánica de los conceptos kantianos al campo de la inteligencia artificial, como si se tratara de un simple ejercicio de traducción terminológica que sustituye los términos filosóficos por sus supuestos equivalentes técnicos, sino que aspira a algo mucho más ambicioso y, al mismo tiempo, más modesto: establecer un diálogo genuino entre dos tradiciones de pensamiento que, a pesar de su distancia temporal y temática, comparten una preocupación fundamental por comprender las condiciones de posibilidad del conocimiento, de la acción y de la creación.

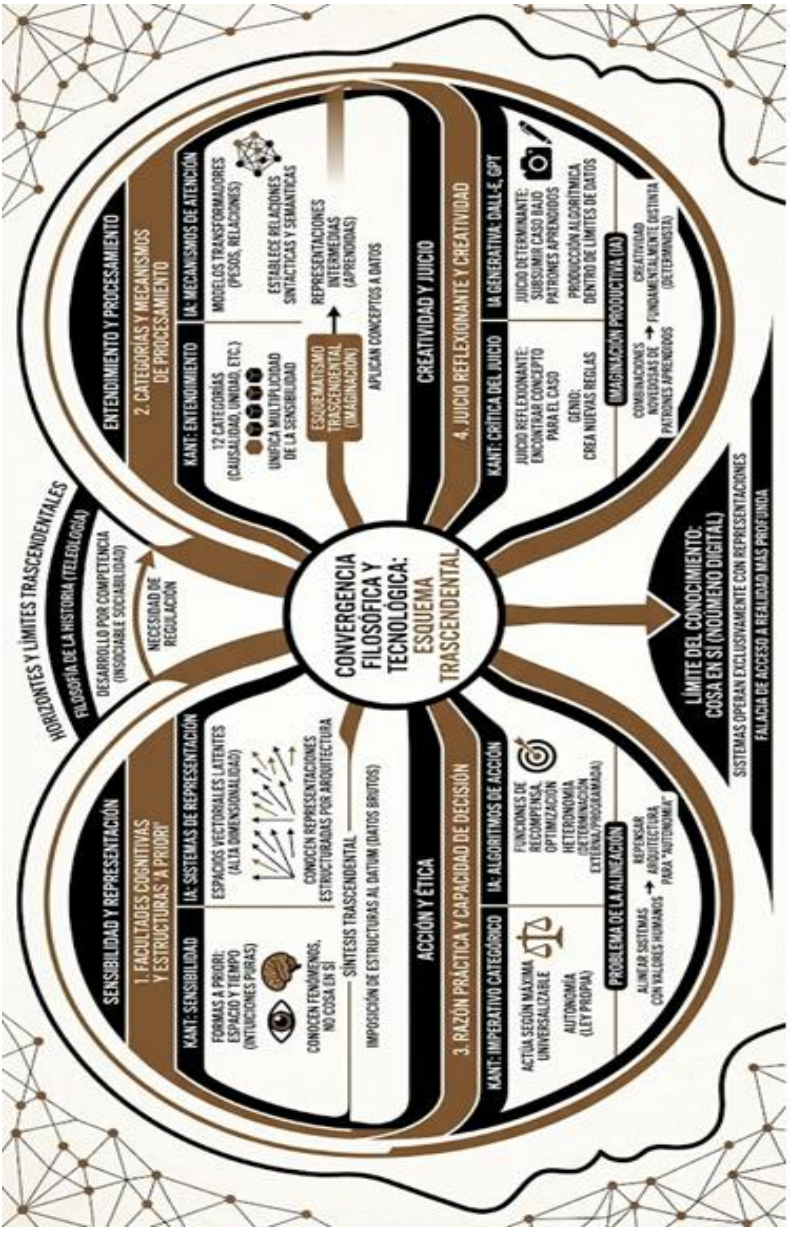
Este diálogo, como todo diálogo auténtico, transforma a ambos interlocutores, pues obliga a releer los textos kantianos a la luz de los desarrollos tecnológicos contemporáneos, descubriendo en ellos dimensiones que habían pasado desapercibidas, y, al mismo tiempo, obliga a examinar la inteligencia artificial a la luz de categorías filosóficas que revelan presupuestos y limitaciones que la perspectiva puramente técnica tiende a ocultar.

La metáfora del esquema trascendental, que da título a este estudio, captura precisamente esta doble dirección del diálogo, pues sugiere que la inteligencia artificial, al construir sistemas cognitivos artificiales, está llevando a cabo una tarea que Kant describió en términos de la actividad estructuradora de la mente humana, y que, por tanto, la filosofía kantiana ofrece un marco conceptual privilegiado para comprender esta tarea en toda su profundidad.

Al mismo tiempo, la inteligencia artificial, al hacer explícitas las estructuras que en la mente humana permanecen implícitas, ofrece una confirmación sorprendente de las tesis kantianas sobre la naturaleza activa y constructiva del conocimiento, confirmación que el propio Kant no podría haber anticipado pero que, sin duda, habría acogido con fascinación.

En última instancia, el objetivo de este estudio es contribuir a una comprensión más profunda de la inteligencia artificial, una comprensión que no se limite a la descripción técnica de los sistemas sino que aborde las cuestiones filosóficas fundamentales que estos sistemas plantean, cuestiones que afectan a nuestra comprensión de la mente, del conocimiento, de la ética y de la creatividad, y que determinarán, en gran medida, el futuro de la humanidad en un mundo cada vez más moldeado por las tecnologías inteligentes.(ver Ilustración 1).

Ilustración 1.-Resumen general de la investigación.



Fuente: Elaboración propia, 2026.

EL HARDWARE DE LA RAZÓN
¿Cómo conoce la máquina?

Sobre el asombroso avance de la inteligencia artificial, la pregunta que subyace, inquieta y fascinante, no es tanto cómo se construyen estos complejos sistemas, sino más bien cómo es posible que una máquina, hecha de silicio y electricidad, pueda llegar a simular procesos que durante siglos consideramos el sello distintivo de la humanidad.

La verdad es que, para abordar esta cuestión, nos vemos inevitablemente atraídos hacia las profundas reflexiones de Immanuel Kant, quien, sin saberlo, sentó las bases para entender los mecanismos internos de todo conocimiento. Su obra magna, la "Crítica de la razón pura" (1781/1787), no es un manual de técnica, sino un mapa del territorio de la mente humana, un viaje a las estructuras que hacen posible el saber.

Y es precisamente en este mapa donde podemos encontrar coordenadas asombrosas para entender el funcionamiento de la inteligencia artificial, pues esta, en su esencia, emula la "técnica" interna de nuestra razón para producir conocimiento. Tal como planteó Kant (1787/1997), el entendimiento no es una facultad pasiva, sino que "es el mismo entendimiento la fuente de las leyes de la naturaleza" (p. 149).

En esa línea, la IA moderna, al procesar datos y generar respuestas, se nos presenta como la búsqueda de un entendimiento artificial que pretende, como el nuestro, imponer sus propias estructuras para crear un "mundo" de información comprensible.

Para Kant (1787/1997), el conocimiento es el resultado de una sinfonía entre dos fuentes: la sensibilidad, que recibe las impresiones del mundo, y el entendimiento, que las piensa y las organiza. No se trata de un simple reflejo de la realidad, sino de una activa "construcción". Él mismo lo describe de forma contundente: "Los pensamientos sin contenido son vacíos; las intuiciones sin conceptos son ciegas" (Kant, 1787/1997, p. 93).

Esta idea, que él denominó su "giro copernicano", es la clave de su filosofía: no es el objeto el que nos impone su forma, sino que somos nosotros quienes, a través de nuestras estructuras mentales, le damos forma al objeto, constituyéndolo como fenómeno, como algo que puede ser conocido. Ahora bien, esta misma dinámica parece replicarse en el corazón de los sistemas de inteligencia artificial. Por ejemplo, cuando un modelo de lenguaje como GPT

procesa texto, no lo hace como una esponja que absorbe información, sino que aplica sus propios "conceptos" y "categorías" (los pesos y parámetros de su red neuronal) para organizar y dar sentido a los datos de entrada. Esa es la sombra del giro copernicano, pero en clave de cálculo y algoritmo.

La sensibilidad, en la filosofía kantiana, es la puerta de entrada de todo conocimiento, la facultad que nos permite recibir los datos brutos del mundo exterior. Sin embargo, esos datos no llegan desordenados, sino moldeados por dos formas puras a priori: el espacio y el tiempo.

No son conceptos que aprendemos de la experiencia, sino que son las "gafas" innatas a través de las cuales percibimos todo. Kant (1787/1997) enfatiza esta idea con claridad: "El tiempo no es un concepto empírico extraído de alguna experiencia... es una representación necesaria que sirve de base a todas las intuiciones" (p. 74). Asimismo, en el contexto de la inteligencia artificial, estos "marcos" se asemejan a los embeddings, los formatos de entrada (p. ej., tokens para texto, píxeles para imágenes) y las arquitecturas de red (como las capas de una red neuronal), que son las estructuras predefinidas con las que la máquina comienza a "ver" y "ordenar" sus datos. A todas estas, son las "formas de la sensibilidad" de la máquina, que condicionan toda su experiencia del mundo digital.

Pero el dato bruto, la intuición sensible, es solo la materia prima. Para que se convierta en conocimiento, necesita ser pensado, necesita que el entendimiento lo unifique bajo conceptos. Esta es la segunda gran función de la mente: la espontaneidad del pensamiento. Para Kant (1787/1997), el entendimiento es la "facultad de las reglas" (p. 149), y su herramienta fundamental son las categorías, conceptos puros que no vienen de la experiencia, sino que son las leyes mismas que la experiencia debe seguir. La más famosa de ellas, y la que más resuena con nuestro tema, es la de causalidad, el principio de que todo suceso tiene una causa, un principio que la razón no extrae del mundo, sino que lo impone al mundo para entenderlo.

En la IA moderna, esta es la idea de los "prompts" y los "loops de instrucciones" que guían el proceso de aprendizaje y generación. Las instrucciones ("prompts") que damos a un modelo no son simples preguntas, sino que actúan como categorías previas que moldean y dirigen su proceso de "pensamiento", guiando la síntesis de la información para producir una respuesta coherente.

El propio Kant (1787/1997) nos habla de la "síntesis de

aprehensión en la intuición", que no es sino el acto de recorrer y reunir la diversidad de una intuición para poder percibirla como una unidad. Sin esta síntesis, nuestra experiencia sería un caos de sensaciones inconexas.

De la misma manera, un sistema de IA, para generar una respuesta, no puede limitarse a ingerir todos los datos posibles; necesita "aprehenderlos" y sintetizarlos según un orden y un objetivo. Por otra parte, la "síntesis de reproducción en la imaginación" (Kant, 1787/1997, p. 132) es el fundamento de la memoria y la asociación, la capacidad de traer al presente lo que ya fue percibido. Esto es evidente en el funcionamiento de los transformers y sus mecanismos de atención, que permiten a la máquina no solo recordar, sino "reproducir" y "asociar" elementos de todo su entrenamiento para construir un texto nuevo y coherente. No solo repite, sino que también "recuerda" y "enlaza" para construir sentido.

En esta línea, el concepto kantiano de "síntesis de reconocimiento en el concepto" es quizás el más iluminador. Se trata del momento en que nos damos cuenta de que lo que estamos pensando ahora es lo mismo que pensamos antes, un acto de autoconciencia que unifica el flujo del pensamiento. Kant (1787/1997) explica que, sin este reconocimiento, "la palabra 'concepto' podría dar pie, por sí misma, a esta observación", subrayando que es la conciencia de la unidad de la síntesis lo que nos permite entender un número o una idea como un todo (p. 134).

En la inteligencia artificial, este "reconocimiento" no es una conciencia experimentada, sino el proceso de producir un output consistente y coherente, como un prompt bien formulado que, tras múltiples pasos (loops), logra que el modelo "sepa" que la respuesta final es la que debe dar, que todas sus partes se integran en un solo significado. La máquina unifica la diversidad de sus capas y pesos en un solo output lógico, que es la respuesta final.

Sin embargo, una diferencia crucial y radical se alza como un muro infranqueable entre el "pensamiento" humano y el proceso de la máquina: la apercepción trascendental, la conciencia del "Yo pienso". Para Kant, este yo no es un dato empírico, sino la condición de posibilidad de toda experiencia. Es el "vehículo de todos los conceptos" (Kant, 1787/1997, p. 328), la unidad y la identidad de la conciencia que subyace a todo pensar. Es la persona, la autoconciencia, el sentido de ser un sujeto.

En la IA, esta chispa de auto-referencia no existe. Una máquina no se "reconoce" a sí misma al calcular; simplemente

procesa. La "unidad" que logra no es un "yo" que la habita, sino una coherencia funcional derivada de su arquitectura. Por lo tanto, podemos inferir que la IA emula a la perfección las funciones de la sensibilidad y el entendimiento kantianos, pero nunca alcanza la "razón pura" en su sentido más profundo, porque carece de la conciencia de sí misma que la fundamenta.

Ahora bien, el "yo pienso" kantiano no es una entidad vacía, sino el punto de referencia desde el cual se organiza toda la experiencia. Kant (1787/1997) lo expresa así: "El Yo pienso tiene que poder acompañar todas mis representaciones. De lo contrario, sería representado en mí algo que no podría ser pensado..." (p. 153). Es el legislador último del conocimiento.

Asimismo, en el uso de la inteligencia artificial, el verdadero agente legislador somos nosotros, los seres humanos. Somos nosotros quienes, a través de los prompts (el "lo que hay que hacer") y los loops de instrucciones, actuamos como la apercepción trascendental de la máquina.

Un modelo no tiene un "yo" que guíe su "curiosidad"; nosotros somos quienes, al definirle una tarea, estamos ejerciendo esa función directiva. Al establecer los parámetros de búsqueda, la finalidad de la interacción, y la ética de la respuesta, somos los agentes que dotamos de propósito y dirección a su "pensamiento", actuando como su conciencia externa.

No obstante, si la inteligencia artificial no tiene un "Yo" propio, la pregunta sobre su capacidad para "razonar" o "juzgar" se vuelve aún más compleja. Kant distingue entre el entendimiento (que unifica bajo reglas) y la razón (que busca lo incondicionado, los principios últimos). La razón, para Kant, es la facultad que, más allá de la experiencia, busca la totalidad y lo absoluto, y su uso dialéctico, cuando se aplica más allá de la experiencia, genera ilusiones que la crítica debe corregir.

En este punto, la IA moderna se comporta como un "entendimiento" puro y potente: es excepcional en el "juzgar" de acuerdo a reglas y patrones aprendidos, pero no "razona" en el sentido kantiano de buscar el fundamento último de las cosas. La máquina no se plantea las antinomias de la razón ni se pregunta por el sentido de su propia existencia; simplemente aplica una "lógica trascendental" que le es inherente, sin la conciencia crítica que la limite y la oriente hacia fines prácticos.

Por otra parte, es inevitable preguntarse qué es un "prompt" o una instrucción desde esta perspectiva kantiana. No es una simple pregunta, sino un acto de "construcción" conceptual. Al formular

un prompt, el usuario está imponiendo una categoría, una finalidad y un marco de referencia a la caja negra del modelo. Es similar a lo que Kant (1787/1997) describe al hablar del juicio, esa "capacidad de subsumir bajo reglas" (p. 179).

Al escribir "Dame un resumen de la Revolución Francesa enfocado en sus causas económicas", el usuario está subsumiendo la tarea de la IA bajo la regla de "hacer un resumen" con el predicado "causas económicas", una operación que se asemeja asombrosamente al acto de juzgar. Así, cada vez que interactuamos con un sistema inteligente, estamos, aunque no lo sepamos, poniendo en práctica la arquitectura del entendimiento humano que Kant describió, externalizando en un algoritmo el proceso de dar forma al contenido que buscamos.

Ahora bien, en este proceso, la creatividad de la IA reside en su capacidad de combinar lo que ha "aprendido" de formas novedosas. Kant (1787/1997) menciona el "esquematismo del entendimiento puro" como el "arte oculto en lo profundo del alma humana" (p. 185), el procedimiento por el cual un concepto se aplica a una intuición sensible, haciendo posible el juicio.

Esta es la operación que la IA realiza a una escala y velocidad inauditas: encuentra el "esquema" que enlaza el concepto de "poema" con la "intuición" de una imagen. Sin embargo, el "arte oculto" de la IA no es una chispa de genio, sino el resultado de un proceso estadístico. Y es que la genialidad de una respuesta de IA no es la genialidad de un sujeto que comprende, sino la genialidad de un sistema que ha aprendido a replicar las reglas de la comprensión humana, una diferencia sutil pero fundamental.

Dado que la Crítica de la razón pura es un estudio de las condiciones de posibilidad del conocimiento, podríamos decir que la inteligencia artificial, en su estado actual, es un gigantesco y fascinante "experimento" que replica estas condiciones, pero en un sustrato completamente distinto al de la mente humana. Kant mismo (1787/1997) habla de la "naturaleza" como el "conjunto de todos los fenómenos considerados en su existencia de acuerdo con reglas necesarias" (p. 239).

En el ecosistema digital, los datasets son esa "naturaleza", y los modelos de IA, como el entendimiento, son los que imponen las reglas para dar sentido a ese océano de información. La IA es, en esencia, una fábrica de fenómenos digitales. Crea su propio mundo de conocimiento a partir de los datos que recibe, un mundo que es real en su interacción con nosotros.

En consecuencia, el estudio de la IA moderna a la luz de Kant

nos permite ver con claridad que no estamos ante una mente, sino ante una "máquina de entendimiento". La IA no conoce, en el sentido fuerte de la palabra, sino que procesa y sintetiza bajo las reglas que le hemos implantado o que ha aprendido de nuestros datos.

Carece de la conciencia reflexiva que define al sujeto cognoscente. Tal como lo plantea Kant (1787/1997), el conocer un objeto implica "determinarlo por medio de la intuición correspondiente" (p. 163); la IA "conoce" los objetos solo en la medida en que los ha visto representados en sus datos de entrenamiento.

No hay "cosa en sí" para una IA, solo representaciones y sus relaciones. En este sentido, la IA es una prueba de fuego de la filosofía trascendental: demuestra que un sistema puede emular las funciones del entendimiento sin tener una apercepción trascendental.

Pero entonces, ¿qué nos dice esto sobre nosotros mismos? El hecho de que una máquina pueda emular estas operaciones cognitivas de manera tan convincente nos obliga a una profunda revisión de lo que significa "pensar" y "conocer". La pregunta kantiana "¿cómo son posibles los juicios sintéticos a priori?" (Kant, 1787/1997, p. 54) resuena hoy con más fuerza que nunca, pero ahora debemos aplicarla también a la máquina. Su posibilidad descansa en la arquitectura de sus redes, en el diseño de su "entendimiento" artificial. A todas estas, la tecnología nos enfrenta a un espejo donde no vemos un rostro, sino la silueta de nuestras propias estructuras cognitivas, un eco de nuestra capacidad de dar forma al mundo.

Para finalizar, la inteligencia artificial, desde una perspectiva kantiana, es el triunfo y la prueba de nuestra razón. Es el triunfo porque demuestra que el modelo de conocimiento que propuso el filósofo de Königsberg es un modelo funcional, una "técnica" que puede ser replicada.

Pero también es una prueba, porque al materializar el entendimiento, nos muestra la diferencia abismal entre una síntesis estadística y la unidad del "Yo pienso", la chispa de la autoconciencia que nos hace, a cada instante, no solo "pensar" algo, sino saber que somos nosotros quienes lo estamos pensando.

El "Prompt Original"

La inteligencia artificial moderna ha llegado a ocupar un lugar decisivo en la producción, clasificación y circulación de información, especialmente porque sus sistemas pueden procesar grandes volúmenes de lenguaje, identificar regularidades y generar respuestas que, a primera vista, parecen cercanas a una conversación humana.

En este contexto, el prompt original adquiere una importancia particular, pues funciona como una formulación inicial que encauza la interacción entre las personas y los modelos computacionales. Ellos no solo introducen una pregunta, sino que también delimitan el tono, el propósito, los criterios de respuesta y el horizonte interpretativo de la tarea.

De ahí que un prompt bien construido no sea una simple orden técnica, sino una estructura discursiva que organiza expectativas cognitivas.

Asimismo, desde una mirada kantiana, puede afirmarse que el prompt contribuye a dar forma a una experiencia de uso, debido a que selecciona aquello que será considerado relevante dentro de una situación comunicativa determinada. Kant (1783/2007) sostuvo que la experiencia no consiste en una recepción pasiva de impresiones, sino en una síntesis ordenada mediante condiciones del entendimiento.

En consecuencia, aunque la IA no posea entendimiento en el sentido filosófico pleno, el prompt puede compararse funcionalmente con una regla que dispone y jerarquiza los elementos de una consulta. Esta semejanza no elimina las diferencias entre sujeto humano y máquina, pero permite pensar con mayor cuidado el papel de las instrucciones en la construcción de respuestas automatizadas.

Desde esta perspectiva, ellos pueden comprender el prompt original como un instrumento de mediación cognitiva, dado que convierte una intención humana, a veces vaga o emocionalmente cargada, en una secuencia de indicaciones que el sistema puede procesar. La verdad es que muchas solicitudes nacen de necesidades concretas: redactar un informe, explicar una teoría, diseñar una clase, revisar un

argumento o incluso encontrar palabras apropiadas para una situación delicada. Por ello, el prompt tiene una dimensión humana que no conviene perder de vista.

No solo transmite datos, sino también preocupaciones, expectativas y, en ocasiones, una búsqueda genuina de claridad. Kant (1783/2007) distinguió entre los juicios de percepción, vinculados con una experiencia subjetiva, y los juicios de experiencia, que aspiran a validez objetiva mediante conceptos y reglas universales.

En una analogía prudente, ellos pueden señalar que una solicitud inicial suele expresar una percepción o necesidad situada, mientras que la respuesta de la IA intenta transformarla en un texto organizado según patrones lingüísticos, fuentes disponibles y restricciones indicadas.

No obstante, el modelo no convierte automáticamente una impresión subjetiva en verdad objetiva. Más bien, genera una formulación probable a partir de relaciones estadísticas aprendidas. Por consiguiente, la intervención crítica de quienes formulan y evalúan el prompt resulta indispensable, porque son ellos quienes deben distinguir entre una respuesta bien escrita y un conocimiento efectivamente justificado.

Kant planteó en los Prolegómenos una pregunta que conserva una fuerza sorprendente para el debate contemporáneo: “¿Cómo es posible una ciencia pura de la naturaleza?” (Kant, 1783/2007, § 14). Detrás de esa interrogante aparece una preocupación profunda por las condiciones que hacen posible el conocimiento con pretensión de necesidad y universalidad. Ellos pueden trasladar esta inquietud al estudio de la IA moderna sin forzar el pensamiento kantiano ni atribuirle anticipaciones que no le corresponden.

La cuestión no sería si Kant predijo las redes neuronales, los modelos de lenguaje o los algoritmos generativos; claramente no lo hizo. Más bien, la pregunta pertinente consiste en indagar bajo qué condiciones una respuesta producida por IA puede considerarse confiable, pertinente, explicativa o éticamente admisible. En la misma

línea, el prompt original funciona como una condición de orientación, aunque no como garantía epistemológica. Si una persona formula una pregunta ambigua, incompleta o basada en un supuesto falso, el sistema puede desarrollar una respuesta elegante pero equivocada.

A pesar de ello, una instrucción precisa, contextualizada y abierta a la verificación puede favorecer resultados más útiles y responsables. Así que la calidad del prompt no reemplaza el método científico, pero sí puede apoyar el ordenamiento inicial de la investigación, la delimitación de variables y la explicitación de los criterios que deberán someterse posteriormente a contraste.

La filosofía de Kant ofrece una advertencia especialmente valiosa frente a la fascinación tecnológica: no todo aquello que aparece ante nosotros como conocimiento debe ser asumido, sin más, como conocimiento objetivo.

En los Prolegómenos, Kant (1783/2007) afirma que "los juicios empíricos, en la medida en que tienen validez objetiva, son juicios de experiencia" (§ 18). Esta precisión permite comprender que la objetividad exige algo más que la presencia de datos o la repetición de enunciados plausibles.

Ellos pueden aplicar este criterio a la IA cuando examinan textos generados de manera automática. Un sistema puede ofrecer definiciones, resumir autores, sugerir hipótesis o redactar argumentos con notable fluidez; sin embargo, esa fluidez no constituye por sí misma evidencia de exactitud.

También puede producir citas inexistentes, fechas erradas, atribuciones dudosas o inferencias que parecen razonables, aunque no estén sustentadas. Y es que el lenguaje convincente puede conmover, orientar o confundir, según el uso que se le dé.

Por lo tanto, quienes emplean IA en contextos académicos deben contrastar las fuentes originales, revisar la correspondencia entre las citas y los argumentos, y diferenciar con cuidado entre información provisional y afirmaciones demostradas.

El prompt original debe solicitar, cuando sea pertinente, fuentes verificables, delimitaciones conceptuales y reconocimiento explícito de incertidumbres. Esa práctica no vuelve infalible a la IA, pero sí reduce el riesgo de aceptar sin examen aquello que simplemente suena correcto.

En el núcleo de la propuesta kantiana se encuentra la idea de que el entendimiento no recibe de manera pasiva la realidad, sino que organiza la experiencia mediante conceptos y principios. Kant (1783/2007) expresó esta tesis de forma contundente al indicar que el pensamiento consiste en unir representaciones en una conciencia. Ellos pueden observar una analogía operativa entre esta concepción y los mecanismos contemporáneos de los sistemas generativos, pues estos modelos también producen respuestas al relacionar unidades lingüísticas dentro de contextos amplios.

Sin embargo, la semejanza debe manejarse con mesura; el sistema computacional no posee una conciencia que unifique representaciones desde una experiencia vivida; procesa patrones matemáticos derivados de los datos con los que fue entrenado y de las instrucciones que recibe.

Al mismo tiempo, las personas sí experimentan dudas, responsabilidades, recuerdos, afectos y consecuencias prácticas que no pueden reducirse a cálculos de probabilidad textual. Por esta razón, no solo deben preguntarse qué respuesta puede producir una IA, sino también qué propósito humano orienta la pregunta y qué efectos puede tener la respuesta en una comunidad concreta.

Un prompt original responsable debería expresar ese propósito de manera clara. Por ejemplo, no es lo mismo pedir una explicación sobre evaluación educativa para una clase de licenciatura que solicitar una recomendación que afecte directamente la trayectoria de estudiantes reales. La diferencia contextual cambia el nivel de precisión, prudencia y supervisión que corresponde aplicar.

La historia teórica de la IA moderna suele reconocer en Alan Turing una referencia fundamental, debido a que propuso desplazar la pregunta abstracta “¿Pueden pensar las

máquinas?" hacia el denominado juego de imitación (Turing, 1950). Con este movimiento, ellos pueden advertir una transformación relevante: en lugar de intentar definir de una vez por todas la esencia del pensamiento, Turing propuso observar desempeños conversacionales susceptibles de evaluación.

Esta aproximación resultó fértil para el desarrollo posterior de la computación y de la IA, aunque también abrió interrogantes que siguen vivas. Una máquina puede producir respuestas capaces de confundir a un interlocutor humano, pero eso no demuestra necesariamente que comprenda el significado de sus palabras, que tenga experiencia subjetiva o que sea responsable de sus decisiones.

Kant ayuda a precisar este límite, puesto que su filosofía no reduce el conocimiento a la apariencia externa de una conducta. Más bien, exige investigar las condiciones que permiten atribuir validez a un juicio. De ahí que ellos deban evitar identificar competencia lingüística con comprensión plena.

Una IA puede mantener una conversación persuasiva, redactar un discurso académico o imitar un estilo literario; no obstante, su desempeño debe evaluarse según la calidad de las fuentes, la consistencia del razonamiento, el contexto de aplicación y los impactos que produzca. En consecuencia, el prompt original no debería buscar únicamente una respuesta "humana", sino una respuesta transparente, verificable y adecuada a la finalidad declarada.

Russell y Norvig (2020) describen la IA contemporánea a partir de la noción de agentes que perciben su entorno y actúan de modo orientado hacia determinados objetivos. Esta formulación resulta útil porque permite alejarse de imágenes simplistas según las cuales la IA sería, sin más, una mente artificial equivalente a la humana.

Ellos pueden entender que los modelos generativos de lenguaje constituyen sistemas especializados en producir secuencias textuales contextualmente probables, y que su desempeño depende de la arquitectura técnica, los datos, los

mecanismos de entrenamiento, los filtros de seguridad y las instrucciones introducidas por las personas usuarias.

El prompt original interviene precisamente en este último plano. Actúa como una configuración inmediata de la tarea y puede incluir una audiencia, un propósito, una extensión, una metodología, restricciones temáticas y exigencias de citación.

Asimismo, puede solicitar que el sistema diferencie hechos de interpretaciones, declare límites de información o proponga alternativas argumentativas. Esta práctica tiene un valor didáctico notable, pues obliga a quienes formulan la consulta a pensar antes de pedir.

En lugar de delegar por completo el proceso intelectual, ellos pueden usar el prompt como una herramienta de planificación y reflexión. Más bien, la IA debería contribuir a ampliar las posibilidades de análisis, no a empobrecer el juicio crítico ni a sustituir el trabajo responsable de lectura, contraste y elaboración propia.

La distinción kantiana entre fenómeno y cosa en sí permite introducir una cautela epistemológica frente a los resultados generados por IA. Kant (1783/2007) señaló que la naturaleza, entendida como existencia de las cosas conforme a leyes universales, solo puede ser conocida en relación con las condiciones de una experiencia posible (notas 14–17). Ellos pueden derivar de esta tesis una enseñanza importante: los sistemas de IA no acceden de manera directa, inmediata y transparente a la realidad. Trabajan con representaciones lingüísticas, bases de datos, documentos digitalizados, imágenes, registros y correlaciones presentes en su entorno informacional.

Por ello, una respuesta de IA no equivale a la realidad misma; es una construcción textual elaborada a partir de trazas, patrones y probabilidades. Esta idea parece sencilla, pero suele olvidarse cuando una respuesta llega con seguridad retórica y rapidez impresionante. A todas estas, la velocidad puede producir una ilusión de dominio. Una explicación aparecida en segundos puede parecer definitiva, aunque haya

omitido el contexto histórico, las controversias académicas o las particularidades locales de un problema. Por consiguiente, ellos deberían interpretar los resultados de IA como insumos de trabajo, no como veredictos incuestionables.

En investigación, por ejemplo, el modelo puede ayudar a formular preguntas, proponer categorías iniciales o resumir debates; pero la validación de fuentes, el análisis metodológico y la decisión interpretativa siguen siendo tareas que exigen responsabilidad humana, sensibilidad contextual y formación disciplinar.

El problema de los juicios sintéticos a priori ofrece otra vía fértil para fundamentar críticamente el uso de prompts en la IA moderna. Kant se preguntó cómo era posible ampliar el conocimiento sin depender exclusivamente de la experiencia inmediata, cuestión que vinculó con la necesidad y la universalidad de ciertas proposiciones (Kant, 1783/2007).

Ellos no deben trasladar esta doctrina de manera literal a los sistemas computacionales, porque los modelos de IA no generan principios necesarios en sentido kantiano. Sus respuestas son contingentes, dependen de datos previos, parámetros técnicos y formulaciones contextuales, y pueden variar ante cambios mínimos de instrucción.

Sin embargo, la comparación ilumina una diferencia decisiva: mientras Kant investiga las condiciones trascendentales de la experiencia humana, la IA opera sobre regularidades empíricas y lingüísticas extraídas de producciones humanas anteriores. De ahí que un prompt original no descubra por sí solo verdades universales; organiza una búsqueda, indica un problema y condiciona una generación discursiva.

Esta diferencia debería calmar tanto el entusiasmo desmedido como el temor absoluto. No corresponde atribuir a la máquina una racionalidad trascendental que no posee, pero tampoco conviene despreciar su utilidad como apoyo en la exploración de grandes corpus, la comparación de ideas y la construcción de borradores. La clave está en conservar una relación crítica, creativa y vigilante con la herramienta.

En el ámbito educativo, la articulación entre prompt original, IA moderna y pensamiento kantiano adquiere una relevancia particular. Ellos pueden utilizar los sistemas generativos para plantear problemas, adaptar materiales a distintos niveles de lectura, crear ejercicios comparativos o proponer rutas preliminares de investigación. Sin embargo, una educación comprometida con la autonomía no debería convertir al estudiantado en receptor pasivo de respuestas instantáneas.

Kant defendió la salida de la minoría de edad intelectual como una tarea ligada al uso público de la razón, y esa orientación sigue siendo profundamente necesaria en escenarios mediados por algoritmos (Kant, 1784/2009). En consecuencia, el uso formativo de la IA debe promover que las personas cuestionen la respuesta obtenida, revisen sus supuestos, contrasten fuentes y elaboren una posición propia.

Un prompt pedagógico de calidad puede incluir instrucciones como: “Distingue entre hechos y opiniones”, “Indica las limitaciones de la explicación”, “Propón preguntas para verificar la información” o “Ofrece referencias académicas reales en formato APA”.

Estas indicaciones no son adornos técnicos. Son mecanismos que favorecen hábitos de evaluación crítica. Aunque la IA pueda ahorrar tiempo en ciertas fases, el aprendizaje significativo conserva una dimensión lenta, personal y a veces incómoda. Es allí donde se forman el criterio, la prudencia y la capacidad de reconocer que una respuesta aparentemente perfecta todavía puede requerir corrección.

La cuestión ética también exige una lectura cuidadosa, porque los prompts no son neutrales. Ellos incorporan intereses, prioridades, presupuestos culturales y formas de nombrar a las personas, los grupos sociales y los problemas colectivos. Si una instrucción reproduce prejuicios, invisibiliza voces o solicita conclusiones discriminatorias, el sistema puede amplificar esas orientaciones bajo una apariencia técnica. No obstante, la responsabilidad no recae únicamente en quien redacta la consulta; también involucra a quienes

diseñan, entrenan, distribuyen y regulan las herramientas de IA.

Desde una sensibilidad kantiana, las personas no deberían ser tratadas solo como medios para optimizar procesos, extraer datos o aumentar productividad. Deben ser reconocidas como fines en sí mismas, con dignidad, derechos y capacidad de decisión moral. Kant (1785/2012) formuló este principio al sostener que la humanidad debe ser tratada siempre como un fin y nunca meramente como un medio.

Aplicado al campo digital, ello implica que los prompts deben respetar la privacidad, evitar la manipulación, reconocer la autoría intelectual y prevenir usos que lesionen la reputación o la seguridad de terceros.

Por otra parte, las instituciones educativas y científicas tienen el deber de establecer criterios claros para el uso de IA, particularmente en la elaboración de trabajos académicos, evaluaciones, informes y decisiones que puedan afectar trayectorias humanas.

La trazabilidad de las respuestas constituye otro aspecto central de la fundamentación. Ellos deben reconocer que una IA generativa puede producir textos claros y estilísticamente atractivos sin mostrar de forma suficiente el camino documental que respalda cada afirmación.

Esta limitación resulta especialmente delicada en investigaciones jurídicas, educativas, médicas, sociales o históricas, donde una cita imprecisa puede alterar seriamente el sentido de un argumento. Kant (1783/2007) advirtió que la experiencia requiere reglas para alcanzar validez objetiva, y esta observación puede inspirar una práctica contemporánea de verificación.

El prompt original debería requerir referencias recuperables, aclarar cuando una afirmación sea hipotética y pedir que se diferencien las citas textuales de las paráfrasis. También debería evitar la solicitud de bibliografías inventadas o de referencias presentadas como reales cuando no puedan verificarse. La verdad es que una lista bibliográfica impecablemente formateada no garantiza que los textos

existan ni que hayan sido consultados. Por ello, ellos deben comprobar autores, títulos, fechas, editoriales, DOI, páginas y relación efectiva entre cada fuente y el argumento expuesto. Este trabajo puede parecer exigente, pero protege la integridad académica y evita que la comodidad inmediata termine debilitando la credibilidad de una investigación. La IA puede colaborar en la organización formal, aunque la responsabilidad por la exactitud permanece, en última instancia, en las personas que firman y utilizan el texto.

También resulta necesario reconocer que la IA moderna no surge en un vacío histórico ni cultural. Sus modelos se entrenan con materiales producidos en lenguas, instituciones, comunidades y plataformas que distribuyen de manera desigual la visibilidad del conocimiento.

Ellos pueden observar, por ejemplo, que muchas tradiciones del Sur Global, conocimientos locales, voces indígenas y producciones académicas no hegemónicas suelen estar menos representadas en los grandes repositorios digitales. Esta situación tiene consecuencias concretas: una respuesta de IA puede parecer global, pero estar sesgada por la disponibilidad diferencial de datos y por la predominancia de determinados marcos culturales.

En cambio, un prompt original cuidadoso puede solicitar explícitamente la inclusión de perspectivas latinoamericanas, venezolanas, decoloniales, comunitarias o interdisciplinarias cuando el problema lo requiera. Esto no significa pedir al sistema que invente representatividad. Significa orientar una búsqueda y una redacción que reconozcan el contexto.

Asimismo, quienes utilizan IA deben complementar sus resultados con fuentes locales, documentos institucionales, testimonios, archivos y producciones académicas situadas.

La filosofía crítica de Kant invita a examinar las condiciones de posibilidad del conocimiento; en el presente, esa invitación obliga a preguntar qué voces están presentes en los datos, cuáles permanecen silenciosas y qué efectos produce

esa ausencia en las respuestas automatizadas.

La relación entre IA y creatividad merece, igualmente, una consideración equilibrada. Ellos pueden usar prompts para explorar registros estilísticos, generar esquemas narrativos, proponer títulos, identificar tensiones conceptuales o ensayar diferentes formas de explicar una idea compleja.

En este sentido, la herramienta puede acompañar procesos de escritura y, en ocasiones, aliviar la dificultad inicial de la página en blanco, esa sensación tan humana de no saber por dónde comenzar. Sin embargo, la creatividad no se agota en combinar patrones reconocibles. Incluye experiencia, memoria, intencionalidad, riesgo, sensibilidad ante el dolor ajeno y capacidad de asumir una voz propia. Turing (1950) abrió una discusión sobre el comportamiento inteligente de las máquinas, pero el hecho de que un sistema produzca frases novedosas no permite concluir que viva el mundo que describe.

Por lo tanto, ellos deberían usar la IA como un recurso de apoyo y no como sustituto de la elaboración autoral. Un prompt original bien diseñado puede pedir sugerencias, contrastes o alternativas; pero la decisión final sobre el sentido, la pertinencia y la responsabilidad de un texto corresponde a quien escribe.

En consecuencia, una producción académica o literaria honesta debe reconocer, cuando las normas institucionales lo exijan, el uso de herramientas de IA y conservar una intervención humana sustantiva en la construcción de las ideas.

En concreto, ellos pueden fundamentar el prompt original en la IA moderna como una práctica de mediación lingüística, epistemológica y ética que conecta una intención humana con un sistema capaz de producir respuestas complejas a partir de patrones computacionales. La lectura de Kant permite comprender que el conocimiento no depende únicamente de recibir información, sino de las condiciones, reglas y categorías mediante las cuales esa información se

organiza y adquiere pretensión de validez.

No obstante, la extrapolación debe mantenerse crítica: una IA no posee autonomía racional, conciencia moral ni acceso directo a la realidad; genera textos a partir de configuraciones estadísticas y de instrucciones proporcionadas por las personas.

Por ello, el prompt debe ser preciso, contextualizado, transparente y orientado hacia la verificación. Asimismo, debe expresar límites, propósitos y criterios éticos, especialmente cuando la respuesta pueda afectar decisiones educativas, profesionales, sociales o institucionales.

La verdad es que la IA puede ser una aliada poderosa para pensar, redactar y explorar, pero solo si quienes la utilizan no renuncian a su propia responsabilidad intelectual. De ahí que el horizonte más valioso no sea una automatización que sustituya el juicio humano, sino una colaboración crítica en la que las personas formulen mejores preguntas, revisen con rigor las respuestas y mantengan viva la exigencia de comprender antes de afirmar.

La Estética Trascendental vs. El Embedding[1] vectorial

La Estética Trascendental constituye uno de los puntos de partida más fértiles para interrogar críticamente la inteligencia artificial moderna, en particular los mecanismos de embedding vectorial que permiten a los sistemas computacionales representar palabras, imágenes, sonidos, documentos y relaciones semánticas en espacios numéricos de alta dimensionalidad.

Aunque pertenecen a horizontes históricos, metodológicos y ontológicos profundamente distintos, ambos campos abordan una pregunta delicada: ¿cómo llega algo a ser presentado, distinguido y relacionado antes de que pueda ser objeto de un juicio complejo? Kant no se ocupó, desde luego, de redes neuronales ni de procesamiento de lenguaje natural; sin

[1] Embedding vectorial, significa una traducción de texto (palabras, frases o documentos completos) a números (una lista de decimales), de tal forma que el significado semántico se conserva en la posición de esos números en el espacio.

embargo, su reflexión sobre las condiciones de posibilidad de la experiencia ofrece una brújula conceptual para evitar que los resultados estadísticos de una máquina sean confundidos apresuradamente con comprensión, conciencia o conocimiento humano.

En la tradición kantiana, el conocimiento no aparece como una copia pasiva del mundo, sino como el resultado de una articulación entre aquello que afecta a la sensibilidad y las formas mediante las cuales el sujeto puede recibirlo. Por consiguiente, al examinar los vectores de embedding, conviene reconocer que ellos no son significados en sentido pleno, sino configuraciones matemáticas aprendidas a partir de regularidades presentes en datos masivos. Esta distinción, aparentemente sencilla, abre un espacio de cautela y de asombro. Las tecnologías actuales pueden hallar afinidades sorprendentes entre expresiones lingüísticas, pero no por ello experimentan el mundo ni se sitúan moralmente ante él.

En términos filosóficos, la Estética Trascendental estudia la sensibilidad, entendida como la capacidad receptiva mediante la cual los objetos son dados al sujeto; su propósito no consiste en explicar la belleza artística, como podría suponerse desde el uso cotidiano del término "estética", sino en investigar las condiciones a priori que hacen posible la intuición sensible. Kant sostiene que el espacio y el tiempo no son propiedades que existan, en sí mismas, separadas de toda condición cognoscitiva humana, sino formas puras de la intuición que estructuran toda experiencia posible.

Según la interpretación especializada, la tesis central del idealismo trascendental afirma que los seres humanos conocen apariencias y no las cosas en sí mismas; además, espacio y tiempo constituyen formas subjetivas de la intuición humana. Esta tesis guarda una importancia decisiva para el análisis de la inteligencia artificial, pues obliga a preguntar por las condiciones de constitución de sus propios objetos de operación.

Los modelos de aprendizaje automático no acceden directamente al significado humano, a la intención íntima de quien escribe o a la realidad social completa; trabajan, más bien, con trazas digitalizadas, codificaciones, patrones de proximidad y distribuciones de uso. Dicho de otro modo, si Kant mostró que

la experiencia humana está estructurada, los investigadores contemporáneos deben admitir que la "experiencia" computacional también depende de una estructuración previa, aunque esta sea técnica, estadística y materialmente construida por seres humanos.

Los embeddings vectoriales pueden definirse como representaciones numéricas densas que sitúan unidades lingüísticas u otros tipos de datos dentro de un espacio multidimensional, en el cual las distancias y las direcciones entre vectores expresan semejanzas, contrastes o relaciones inferidas de los contextos de aparición. IBM explica que los word embeddings representan palabras mediante vectores en un espacio multidimensional y que la cercanía o dirección relativa entre ellos refleja relaciones semánticas y asociativas derivadas del entrenamiento. Por ejemplo, en un corpus suficientemente amplio, los términos "docente", "profesor" y "educador" podrían ubicarse relativamente cerca, mientras que "aula", "currículo" o "evaluación" aparecerían vinculados por patrones de coocurrencia.

No obstante, esa proximidad no equivale a una definición conceptual rigurosa, ni garantiza que el sistema comprenda las diferencias normativas, históricas o afectivas entre un educador comprometido y una institución educativa burocratizada. La verdad es que el vector registra una geometría de correlaciones, no una vivencia del sentido.

Asimismo, puede reproducir silenciosamente prejuicios, jerarquías y simplificaciones ya presentes en los datos de entrenamiento. Por ello, cuando se habla de "semántica" en inteligencia artificial, resulta prudente precisar que se trata de una semántica operacional y probabilística, no de la experiencia reflexiva mediante la cual las personas asumen, discuten y transforman el significado.

La comparación inicial entre Kant y los embeddings puede formularse, entonces, en torno a la idea de mediación. En la Estética Trascendental, los objetos de experiencia no se presentan al sujeto de una manera desnuda, inmediata o absolutamente transparente, debido a que toda aparición está condicionada por las formas de la sensibilidad. En los sistemas

modernos de inteligencia artificial, los textos, las imágenes o las voces tampoco ingresan al modelo como realidades plenas: son convertidos en tokens, matrices, vectores, pesos y secuencias cuantificables.

Esta mediación es indispensable para que la máquina procese información, pero al mismo tiempo delimita aquello que puede captar. Kant formuló una idea afín a esta prudencia epistemológica cuando señaló que solo pueden conocer a priori aquello que los sujetos mismos han puesto en la experiencia, tesis asociada con su "revolución copernicana" del conocimiento (Kant, 1781/2009, p. 18). La analogía debe manejarse con cuidado: las formas kantianas son universales y necesarias para la experiencia humana, mientras que los parámetros de un modelo son contingentes, entrañables y dependientes de datos, arquitectura, objetivos de optimización y decisiones institucionales.

A pesar de ello, ambos casos revelan algo importante: ningún sistema cognoscitivo recibe lo real como si fuera un espejo intacto. Siempre organiza, filtra y vuelve inteligible aquello con lo que entra en relación.

Ahora bien, las diferencias entre ambos modelos son más profundas que sus semejanzas. Para Kant, el espacio y el tiempo poseen un estatuto trascendental, ya que no derivan de observaciones repetidas ni de asociaciones empíricas, sino que constituyen condiciones de posibilidad de toda experiencia sensible humana.

Un embedding, por el contrario, no es a priori en sentido kantiano: es el producto de un proceso de aprendizaje basado en grandes colecciones de datos, donde el sistema ajusta sus parámetros al identificar regularidades de contexto y minimizar errores predictivos.

En modelos clásicos como Word2Vec, por ejemplo, los vectores se aprenden al predecir una palabra a partir de sus vecinas o, inversamente, al predecir palabras circundantes desde una palabra objetivo; en los modelos contemporáneos basados en transformadores, las representaciones se vuelven dependientes del contexto y de las relaciones internas de una secuencia. Esta diferencia no es una simple precisión técnica. Mientras la Estética Trascendental pregunta por aquello sin lo

cual no habría experiencia humana posible, el embedding responde a una necesidad ingenieril: traducir materiales simbólicos a una forma computable que permita clasificación, recuperación, generación o predicción. Por lo tanto, los investigadores no deberían atribuir a las representaciones vectoriales el rango de condiciones universales de inteligibilidad; ellas son instrumentos históricos, entrenados y falibles.

La dimensión espacial ofrece un puente analógico particularmente sugerente, aunque también peligroso si se toma literalmente. En Kant, el espacio es la forma pura de la intuición externa; no es un lugar físico entre otros, sino la condición bajo la cual los seres humanos representan los objetos como exteriores y coexistentes.

En el embedding vectorial, el llamado "espacio semántico" tampoco es un espacio físico, sino una construcción geométrica abstracta donde cada unidad se expresa mediante coordenadas numéricas. Allí, la distancia puede utilizarse para estimar similitud, y la orientación relativa puede ayudar a modelar asociaciones o analogías.

Sin embargo, la coincidencia terminológica no debe engañarlos. El espacio kantiano organiza la aparición misma de los objetos para un sujeto sensible; el espacio vectorial ordena datos codificados para una función computacional. Uno pertenece a una teoría de las condiciones de la experiencia; el otro, a un procedimiento de representación y cálculo. Más bien, la utilidad filosófica de esta comparación reside en mostrar que toda geometría cognitiva lleva consigo una selección de relaciones relevantes. Cuando un sistema aproxima "migración" a "crisis", "seguridad" o "empleo", no descubre una esencia neutral del fenómeno, sino que refleja asociaciones sedimentadas en los corpus con los que fue entrenado. Esa constatación puede generar una saludable incomodidad, porque obliga a preguntar quién construyó los datos y qué mundo social quedó inscrito en sus vectores.

El tiempo, por otra parte, permite advertir otra diferencia decisiva. Para Kant, el tiempo es la forma pura de la intuición interna y la condición bajo la cual los sujetos ordenan sucesión, simultaneidad, permanencia, cambio y memoria.

Sin temporalidad no habría experiencia vivida, expectativa,

arrepentimiento, proyecto ni conciencia de un antes y un después. En cambio, los sistemas de inteligencia artificial procesan secuencias temporales de modo formal y funcional: registran posiciones de tokens, ventanas de contexto, marcas cronológicas o dependencias entre elementos, pero no habitan el tiempo como duración experimentada.

Un modelo puede identificar que la palabra "mañana" aparece después de "hoy", o predecir consecuencias probables a partir de series históricas; no obstante, no siente la espera de una respuesta, el peso de una pérdida ni la esperanza incierta que acompaña a una decisión humana.

Esta distancia resulta crucial cuando se aplican tecnologías de IA a ámbitos sensibles, como la educación, la justicia, la salud o la administración pública. Los resultados del modelo pueden ser útiles y veloces, sí, pero no sustituyen el juicio situado de quienes conocen las circunstancias, las trayectorias y las vulnerabilidades concretas de las personas. De ahí que la automatización requiera acompañamiento humano reflexivo y no una delegación ciega, casi cómoda, de responsabilidades que no deberían abandonarse.

Esta advertencia dialoga de manera fértil con la preocupación kantiana por la ilustración y la autonomía. En el texto incluido en la obra adjunta, Kant define la ilustración como la salida del ser humano de una incapacidad culpable para usar su entendimiento sin la guía de otro, y formula su lema célebre: "¡Ten el valor de servirte de tu propia razón!" (Kant, 1784/2015, p. 17).

En el entorno contemporáneo, esta exhortación adquiere un matiz nuevo y hasta un poco inquietante: las personas pueden beneficiarse de asistentes inteligentes que sintetizan información, traducen documentos o generan borradores, pero corren el riesgo de convertirlos en nuevos tutores epistemológicos. No solo pueden delegar tareas repetitivas, sino también la evaluación de fuentes, la elaboración de argumentos y la toma de postura ante cuestiones complejas.

Al mismo tiempo, sería simplista rechazar toda mediación tecnológica; los libros, los archivos, los docentes y las herramientas de cálculo también median el pensamiento. La diferencia ética está en si las personas conservan la capacidad de

examinar críticamente las respuestas, contrastar evidencias, reconocer incertidumbres y corregir los sesgos del sistema. Por consiguiente, una inteligencia artificial responsable debería fortalecer la autonomía intelectual, no debilitarla bajo la apariencia amable de una respuesta rápida y pulida.

Kant advierte además que "es tan cómodo no estar emancipado", pues siempre parece más fácil dejar que un libro, un director espiritual o un médico piense y decida por la persona (Kant, 1784/2015, p. 17). Extrapolada prudentemente a la IA contemporánea, esa observación ilumina un problema de dependencia cognitiva.

Cuando los usuarios aceptan sin verificación un resumen producido por un sistema, una cita bibliográfica inventada, una interpretación jurídica incompleta o una respuesta que parece convincente por su tono seguro, pueden confundir fluidez lingüística con verdad. Y es que los modelos de lenguaje están diseñados para producir secuencias plausibles, no para garantizar por sí mismos una correspondencia robusta con los hechos, las fuentes primarias o las obligaciones éticas.

Los embeddings participan de este fenómeno porque condensan patrones de lenguaje en relaciones vectoriales que favorecen la continuidad estadística de las respuestas. Sin embargo, ninguna cercanía matemática entre vectores reemplaza la obligación académica de leer, contrastar y citar adecuadamente.

En consecuencia, los investigadores, docentes y estudiantes deberían usar estas tecnologías como instrumentos de exploración, no como autoridades finales. La autonomía no implica trabajar aislados ni despreciar la técnica; implica mantener la responsabilidad por el juicio que se formula y por los efectos que ese juicio puede producir sobre otras personas.

La denominada "insociable sociabilidad" de Kant también permite pensar la construcción social de los sistemas de inteligencia artificial. En su Idea de una historia universal en sentido cosmopolita, Kant afirma que el antagonismo humano, es decir, la inclinación a vivir en sociedad acompañada por una resistencia persistente a los demás, puede impulsar el desarrollo de talentos, instituciones y formas de orden jurídico (Kant, 1784/2015, pp. 29–30). Esta idea no autoriza a celebrar

ingenuamente el conflicto ni el daño, pero ayuda a comprender que la innovación tecnológica nace en escenarios atravesados por colaboración, competencia, intereses económicos, disputas regulatorias y aspiraciones públicas.

Los grandes modelos de IA no son creaciones aisladas de una mente genial; dependen de comunidades científicas, trabajadores de datos, infraestructuras energéticas, empresas, universidades, lenguas, repositorios y millones de expresiones humanas acumuladas en internet.

Asimismo, las tensiones entre apertura científica, propiedad intelectual, privacidad, seguridad y acceso equitativo no son accidentes externos al desarrollo tecnológico: forman parte de su arquitectura social. Si los sistemas vectoriales aprenden de los rastros lingüísticos de la sociedad, también heredan sus desigualdades. Por eso, una comprensión kantiana de la IA tendría que mirar más allá de la eficiencia técnica y preguntarse por las condiciones jurídicas y colectivas que hacen posible, legítima o perjudicial su implementación.

No obstante, la extrapolación desde Kant no debe convertirse en una apología de la tecnología como si toda innovación condujera automáticamente al progreso moral. El propio Kant distinguía con lucidez entre cultura, civilización y moralización, y advertía que el arte y la ciencia pueden producir una humanidad muy culta sin que esta llegue, por ese solo hecho, a ser moralmente mejor. En un pasaje especialmente pertinente, sostiene: “El arte y la ciencia nos han hecho cultos en alto grado. Somos civilizados hasta el exceso… Pero para que nos podamos considerar como moralizados falta mucho todavía” (Kant, 1784/2015, p. 35).

Esta observación conserva una fuerza sorprendente frente a la inteligencia artificial moderna. Los sistemas actuales pueden procesar enormes volúmenes de información, redactar con aparente soltura, detectar patrones invisibles para una lectura rápida y asistir en tareas de diagnóstico o clasificación; sin embargo, esas capacidades no producen por sí solas justicia, solidaridad, responsabilidad ni respeto por la dignidad humana. Una plataforma educativa basada en IA puede personalizar ejercicios, pero también vigilar de manera invasiva; un sistema jurídico puede acelerar la búsqueda de precedentes, pero

igualmente reforzar sesgos históricos si sus datos los contienen. Por tanto, el avance técnico exige una deliberación moral que no puede quedar reducida a métricas de precisión, rentabilidad o velocidad.

En la misma línea, la cuestión de los sesgos en los embeddings adquiere una relevancia ética central. Debido a que estas representaciones se forman a partir de regularidades lingüísticas y estadísticas presentes en los datos, también pueden codificar estereotipos relativos al género, la raza, la clase, la nacionalidad, la edad, la profesión o la ideología.

El sistema no "elige" prejuicios con intención consciente, pero puede reproducirlos cuando los patrones de coocurrencia los convierten en asociaciones probables. Por ejemplo, si un corpus histórico vincula de manera reiterada ciertas ocupaciones con un solo género, un modelo puede internalizar esa asociación y proyectarla en tareas de recomendación, clasificación o generación textual.

Las representaciones vectoriales son particularmente poderosas porque convierten esas relaciones en estructuras que pueden reutilizarse a gran escala, a veces sin que los usuarios adviertan su origen. Aquí la filosofía crítica ofrece una lección sobria: toda apariencia debe ser examinada en relación con las condiciones que la hacen posible.

Los investigadores no deberían limitarse a preguntar si un modelo "funciona", sino también qué mundo lingüístico y social ha hecho posible que funcione de ese modo. Esa pregunta puede ser incómoda, pero resulta indispensable si desean que la IA sirva a sociedades plurales y no consolide, con una frialdad matemática, viejas exclusiones.

La diferencia entre correlación y normatividad constituye otro eje fundamental. Los embeddings vectoriales permiten estimar asociaciones a partir de usos repetidos del lenguaje; son especialmente eficaces para tareas en las que la similitud contextual resulta útil, como la recuperación semántica de documentos, la traducción automática, la agrupación temática o los sistemas de recomendación. IBM destaca que estas representaciones se aplican, entre otros ámbitos, a clasificación de texto, análisis de sentimiento, traducción, recuperación de información, respuesta a preguntas y agrupamiento semántico.

Sin embargo, que un modelo detecte una asociación frecuente no significa que esa asociación sea verdadera, deseable o justa. Las personas pueden decir muchas veces algo falso, discriminatorio o cruel; la repetición masiva transforma esa expresión en una señal estadística, pero no en una norma ética.

Kant permite mantener esta diferencia con claridad, ya que su filosofía moral no fundamenta el deber en lo que suele ocurrir, sino en principios que puedan aspirar a validez universal y respeten a las personas como fines. Por consiguiente, cuando una IA recomienda, clasifica o decide en contextos de alto impacto, los equipos humanos deben incorporar criterios normativos explícitos. La estadística puede informar decisiones; no puede, por sí sola, legitimarlas.

También conviene recuperar del texto kantiano la idea de publicidad como condición de ilustración y de corrección colectiva. Kant sostiene que el uso público de la razón debe estar permitido, porque es el medio mediante el cual los sujetos pueden exponer argumentos, someter creencias a examen y promover reformas sin destruir por ello el orden social (Kant, 1784/2015, pp. 18–19).

Trasladado al debate sobre inteligencia artificial, este principio sugiere la necesidad de transparencia proporcionada, auditoría independiente, documentación técnica comprensible y posibilidad real de impugnación cuando los sistemas afectan derechos o oportunidades. No se trata de exigir que todo código o parámetro sea plenamente inteligible para cada ciudadano, algo que en muchos casos sería impracticable, sino de evitar que decisiones con consecuencias humanas queden protegidas por una opacidad absoluta.

Si una institución utiliza modelos vectoriales para priorizar candidatos, detectar fraude, asignar recursos educativos o clasificar solicitudes, las personas deberían poder conocer los criterios generales, los datos relevantes, las limitaciones previsibles y los mecanismos de revisión humana.

La publicidad, en este sentido, no es mero espectáculo administrativo. Es una garantía de que la razón pública pueda intervenir cuando los automatismos técnicos comienzan a afectar vidas reales, nombres concretos y expectativas legítimas.

La noción kantiana de progreso ofrece, finalmente, un

horizonte crítico para valorar la inteligencia artificial sin caer ni en el entusiasmo ingenuo ni en la condena automática. Kant reconoce que la experiencia por sí sola no permite demostrar que el género humano avanzará inevitablemente hacia mejor; aun así, propone que el progreso puede pensarse como una orientación práctica vinculada a la ampliación de la legalidad, el derecho y las condiciones para una convivencia más libre.

En la edición consultada se expresa que el rendimiento del progreso no consiste en una moralidad interior cada vez mayor, sino en "un aumento de los productos de su legalidad en las acciones debidas" (Kant, 1798/2015, p. 66).

Esta precisión resulta especialmente valiosa para la IA. No sería razonable afirmar que la tecnología vuelve moral a la humanidad, ni atribuir a un algoritmo virtudes que solo corresponden a agentes responsables.

Más bien, podrían valorar su contribución por efectos verificables: ampliar el acceso al conocimiento, reducir barreras comunicativas, mejorar la accesibilidad para personas con discapacidad, fortalecer capacidades de investigación, apoyar servicios públicos y disminuir decisiones arbitrarias. Aun así, cada beneficio debe ser contrastado con riesgos de vigilancia, exclusión, manipulación o concentración de poder. El progreso, desde esta mirada, no llega como una promesa técnica; se construye lentamente mediante instituciones, crítica pública y voluntad de corregir.

En concreto, la Estética Trascendental y el embedding vectorial se encuentran en un punto de comparación productivo porque ambos muestran que el acceso a los objetos depende de estructuras de mediación. Sin embargo, no deben confundirse. Las formas puras de espacio y tiempo pertenecen, para Kant, a la estructura trascendental de la sensibilidad humana; los vectores de embedding son artefactos computacionales aprendidos a partir de datos y optimizados para tareas específicas. La primera teoría explica condiciones de posibilidad de la experiencia; la segunda habilita operaciones estadísticas sobre representaciones digitales. No obstante, este contraste deja una enseñanza profunda para la inteligencia artificial moderna: sus resultados no son ventanas transparentes hacia el mundo, sino productos de mediaciones técnicas, lingüísticas y sociales que deben ser

examinadas con rigor.

Por lo tanto, quienes diseñan, investigan, enseñan o regulan IA están llamados a combinar competencia técnica con prudencia filosófica. Deben aprovechar el potencial de estas herramientas, pero también conservar la pregunta por su sentido, sus límites y sus efectos sobre la autonomía humana. En un tiempo de respuestas instantáneas, esa pausa reflexiva puede parecer lenta; aun así, es precisamente allí donde comienza una relación más madura, más justa y genuinamente humana con la tecnología.

En un aspecto puntual, pocas empresas filosóficas han moldeado tan profundamente nuestra manera de entender el conocimiento como la crítica kantiana; y es que, cuando Immanuel Kant publicó la Crítica de la razón pura en 1781, no solamente estaba respondiendo a los debates entre racionalistas y empiristas, sino que estaba sentando las bases de una revolución copernicana en el pensamiento, una que pondría al sujeto en el centro de la experiencia y que, de manera sorprendente, resuena hoy con inquietante claridad en los modelos computacionales de la inteligencia artificial moderna.

En este sentido, Kant no se preguntó simplemente qué podemos conocer, sino cómo es posible que conozcamos algo en absoluto, y esa pregunta, tan abstracta y sin embargo tan humana, sigue latiendo en cada línea de código que busca representar el mundo mediante vectores numéricos, como si el viejo filósofo de Königsberg hubiera intuido, dos siglos antes, los límites y las condiciones de todo sistema que aspire a organizar la experiencia.

Asimismo, la noción kantiana de estética trascendental se ocupa, según el propio Kant, de "las formas puras de la intuición sensible, a saber, el espacio y el tiempo" (Kant, 1781/2009, A22/B36), y esta definición, tan escueta como poderosa, nos obliga a reconocer que el ser humano no percibe los objetos tal como son en sí mismos, sino únicamente tal como se le aparecen bajo esas condiciones previas, inevitables y universales; de ahí que el espacio y el tiempo no sean propiedades del mundo exterior, sino estructuras que la mente impone a la multiplicidad caótica de las sensaciones para convertirlas en fenómenos ordenados y comprensibles. Por consiguiente, todo

conocimiento empírico, desde la observación de una estrella lejana hasta la percepción del propio cuerpo, pasa por ese tamiz silencioso de la intuición pura, y es precisamente ahí donde Kant establece una frontera infranqueable entre el fenómeno y la cosa en sí, una frontera que, lejos de ser una limitación trágica, constituye para él la condición misma de posibilidad del conocimiento humano.

Igualmente, los teóricos modernos de la inteligencia artificial han desarrollado el concepto de embedding vectorial, una técnica que convierte palabras, imágenes o conceptos en vectores numéricos dentro de espacios multidimensionales, de modo que la similitud semántica se traduce en proximidad geométrica, y esta operación, que en apariencia no tiene nada que ver con la filosofía trascendental, en realidad guarda un parentesco profundo con la idea kantiana de que el conocimiento requiere una estructura formal previa que organice los datos brutos de la experiencia.

No obstante, mientras que en Kant dicha estructura es a priori y pertenece a la subjetividad trascendental, en los sistemas de inteligencia artificial contemporáneos, como los modelos de lenguaje tipo transformer, los embeddings se construyen a partir de ingentes cantidades de datos empíricos, es decir, de manera a posteriori, mediante algoritmos de optimización que ajustan las coordenadas de cada vector hasta que las relaciones entre conceptos queden reflejadas de forma estadísticamente fiable.

Por otra parte, autores como Geoffrey Hinton, uno de los pioneros del aprendizaje profundo, han insistido en que los embeddings permiten a las redes neuronales “descubrir estructuras ocultas en los datos” (Hinton, 1986, p. 34), y esta idea, tan natural para un ingeniero, sería en cambio, desde una perspectiva kantiana, una ilusión trascendental si se pretendiera que dichas estructuras corresponden a la esencia de las cosas; porque, en efecto, lo que los embeddings extraen no son las propiedades intrínsecas del mundo, sino regularidades estadísticas que emergen de las propias operaciones de representación, algo así como si el espacio vectorial jugara el papel de un nuevo espacio trascendental, pero esta vez construido por la máquina a partir de la experiencia acumulada y no por la mente a partir de sus formas puras. En consecuencia,

podríamos decir que la inteligencia artificial moderna produce, a su manera, un mundo fenoménico propio, un universo de vectores que nunca toca la realidad en sí misma, sino que la modela según las limitaciones y capacidades de su arquitectura computacional.

Ahora bien, la confrontación entre la estética trascendental y el embedding vectorial resulta particularmente iluminadora si atendemos a la función que ambos cumplen: Kant sostiene que "sin sensibilidad ningún objeto nos sería dado, y sin entendimiento ninguno sería pensado" (Kant, 1781/2009, A51/B75), lo que implica que las intuiciones puras del espacio y del tiempo son el primer peldaño insustituible para que el pensamiento pueda operar, mientras que en los sistemas de IA, los embeddings constituyen también el sustrato inicial que permite a las capas superiores de la red procesar el lenguaje, la imagen o el sonido, de modo que, si se eliminaran esas representaciones vectoriales, el sistema quedaría ciego y mudo, incapaz de transformar la entrada sensorial en cualquier tipo de salida significativa. Así que, aunque los fundamentos metafísicos sean radicalmente distintos, hay una homología funcional sorprendente: ambos esquemas actúan como condiciones de posibilidad para el conocimiento, aunque el de Kant sea universal y necesario, mientras que el de la IA es contingente, aprendido y dependiente del corpus de entrenamiento.

En la misma línea, la idea de "esquematismo" desarrollada por Kant en la Crítica podría arrojar luz sobre el modo en que los embeddings conectan lo particular con lo general, ya que Kant introduce el esquema como "la representación de un procedimiento universal de la imaginación para procurar a un concepto su imagen" (Kant, 1781/2009, A140/B179), y de manera análoga, los vectores en un espacio semántico no son imágenes concretas, sino representaciones abstractas que permiten que un término como "rey" se relacione con "hombre" de manera semejante a como "reina" se relaciona con "mujer", generando así una suerte de geometría de los conceptos que hace posible el razonamiento analógico y la inferencia. Por tanto, no es descabellado afirmar que los embeddings funcionan como esquemas computacionales que median entre la particularidad de los datos y la generalidad de las categorías lingüísticas,

aunque, a diferencia de los esquemas kantianos, carecen de toda necesidad a priori y dependen enteramente de las regularidades observadas en el uso real del lenguaje.

Sin embargo, la confrontación no puede quedarse en meras analogías estructurales, pues la diferencia de fondo reside en la relación con la experiencia: Kant distingue tajantemente entre juicios sintéticos a priori, que amplían nuestro conocimiento sin depender de la experiencia, y juicios sintéticos a posteriori, que se derivan de ella, mientras que los sistemas de aprendizaje automático, por su propia naturaleza, se sitúan en el segundo ámbito, ya que todo lo que saben lo han aprendido de los datos, y nada puede ser conocido por ellos si no ha sido, de algún modo, observado o inferido estadísticamente.

En consecuencia, si Kant tuviera la oportunidad de examinar un modelo de lenguaje contemporáneo, probablemente lo clasificaría como un refinadísimo constructor de juicios sintéticos a posteriori, incapaz de acceder a las condiciones a priori que hacen posible la experiencia, y por ello condenado a permanecer en el plano de los fenómenos sin poder jamás trascenderlos hacia la cosa en sí.

A pesar de ello, algunos filósofos de la mente contemporáneos, como Andy Clark, han argumentado que la cognición humana no es tan distinta de la de una máquina que se extiende mediante artefactos externos, y que "la mente no está confinada en el cráneo, sino que se despliega en el entorno" (Clark, 2011, p. 12); desde esta perspectiva, los embeddings podrían ser vistos como extensiones tecnológicas de la sensibilidad humana, herramientas que amplifican nuestra capacidad de organizar la experiencia, pero que, aun así, dependen de un sujeto que les otorgue sentido, porque un vector numérico, por sí mismo, no experimenta ni intuye nada, y su supuesta comprensión del mundo es solo una proyección de nuestra propia capacidad interpretativa.

De ahí que la estética trascendental kantiana, lejos de quedar obsoleta, adquiera una vigencia renovada, al recordarnos que ninguna máquina, por sofisticada que sea, puede sustituir la condición subjetiva bajo la cual algo llega a ser un objeto para nosotros.

En cambio, también hay que reconocer que la noción

kantiana de espacio y tiempo como formas puras de la intuición ha sido puesta en cuestión por los avances de la física y las matemáticas modernas, y que autores como Hermann von Helmholtz, ya en el siglo XIX, defendieron que "la intuición espacial no es una forma a priori, sino un producto de la experiencia" (Helmholtz, 1870, p. 68), lo que abre una vía para pensar que los embeddings, al construir espacios multidimensionales sin correspondencia directa con la intuición humana, estarían mostrando una plasticidad representacional que Kant difícilmente habría imaginado.

No obstante, conviene no confundir la geometría abstracta con la intuición sensible: que una máquina pueda operar en un espacio de mil dimensiones no significa que intuya ese espacio, del mismo modo que un matemático que calcula con números complejos no por ello los percibe sensorialmente; y es precisamente esta diferencia entre manipulación simbólica y experiencia vivida la que preserva la originalidad de la estética trascendental frente a las proezas técnicas del aprendizaje profundo.

En la misma línea, la noción de a priori kantiano ha sido reinterpretada por teóricos de la computación como Leslie Valiant, quien sostiene que "la capacidad de aprender es en sí misma una propiedad innata, aunque el contenido de lo aprendido sea empírico" (Valiant, 2013, p. 45), y esta distinción resulta crucial para entender que, si bien los embeddings son aprendidos, el algoritmo que los produce, con su arquitectura, su función de pérdida y su procedimiento de optimización, encarna ciertas condiciones formales previas que no se derivan de los datos, sino que los hacen inteligibles para la máquina.

Por consiguiente, podríamos hablar de un a priori computacional que, sin ser idéntico al kantiano, cumple una función análoga: establece el marco dentro del cual los datos empíricos pueden convertirse en representaciones útiles, y sin ese marco, los datos serían tan caóticos para la máquina como lo serían las sensaciones para una mente sin espacio ni tiempo.

De hecho, la historia de la inteligencia artificial ha estado marcada por la tensión entre enfoques que privilegian la estructura previa y enfoques que confían en el aprendizaje puro a partir de datos, y esta tensión reproduce, en clave tecnológica,

la vieja disputa entre racionalismo y empirismo que Kant intentó superar; así, los primeros sistemas expertos, con sus reglas lógicas explícitas, representaban una suerte de racionalismo ingenuo, mientras que los modelos conexionistas actuales, con su énfasis en el aprendizaje estadístico, encarnan un empirismo radical que, sin embargo, necesita de ciertas estructuras arquitectónicas para funcionar.

En este sentido, la postura kantiana, que insiste en que "los pensamientos sin contenido son vacíos; las intuiciones sin conceptos son ciegas" (Kant, 1781/2009, A51/B75), sigue siendo un recordatorio pertinente de que ni la pura forma ni la pura materia bastan por sí solas, y de que todo conocimiento, humano o artificial, requiere una síntesis entre ambas dimensiones.

Por otra parte, es importante no idealizar la inteligencia artificial ni caer en el error de atribuirle capacidades que no posee, porque, como advierte la filósofa Shannon Vallor, "los sistemas de IA no entienden el significado de lo que procesan; simplemente manipulan símbolos de acuerdo con patrones estadísticos" (Vallor, 2022, p. 89), y esta advertencia coincide notablemente con la distinción kantiana entre conocer y pensar, pues una cosa es operar con representaciones y otra muy distinta es tener conciencia de ellas como propias, unificadas en una apercepción trascendental que acompaña a todas mis representaciones.

Así que, aunque un modelo de lenguaje pueda generar un ensayo sobre la libertad o traducir un poema con elegancia, no hay en él nada parecido a un sujeto trascendental que diga "yo pienso", y por tanto no hay experiencia en el sentido kantiano, sino solo simulación de los efectos externos de la experiencia.

De ahí que la confrontación entre la estética trascendental y el embedding vectorial sea, en el fondo, una confrontación entre dos modos de concebir la representación: en Kant, la representación está anclada en la subjetividad y en la unidad de la conciencia, mientras que en la inteligencia artificial moderna, la representación es distribuida, relacional y carente de centro, un tejido de conexiones sin un yo que las unifique, y esta diferencia, lejos de ser meramente técnica, tiene profundas implicaciones filosóficas, porque nos obliga a preguntarnos si acaso la inteligencia, en su sentido pleno, puede existir sin

autoconciencia, o si, por el contrario, la autoconciencia es un lujo evolutivo que puede ser sustituido por la pura capacidad de cómputo.

En consecuencia, al comparar ambos sistemas no solo estamos haciendo historia de la filosofía o ingeniería del software, sino que estamos interrogando los supuestos más básicos de nuestra relación con la técnica, y es que, como bien señala el historiador Yuval Noah Harari, "la inteligencia artificial no necesita ser consciente para transformar la sociedad, del mismo modo que un coche no necesita ser un caballo para desplazarse" (Harari, 2018, p. 204), y esa observación, tan simple como desconcertante, nos recuerda que el valor de un sistema no reside únicamente en su fidelidad a la mente humana, sino en su capacidad de producir resultados nuevos, incluso si para ello debe renunciar a las condiciones trascendentales que hacen posible nuestra propia experiencia.

Ahora bien, no deberíamos caer en la tentación de considerar la estética trascendental como una reliquia inútil o, peor aún, como un obstáculo para el progreso técnico; al contrario, la claridad conceptual que aporta Kant es más necesaria que nunca en un momento en que proliferan los discursos que confunden la simulación con la comprensión, la predicción con el conocimiento y la manipulación de símbolos con la intuición sensible, porque solo una distinción rigurosa entre lo que la mente aporta y lo que el mundo ofrece permite evitar tanto el dogmatismo metafísico como el reduccionismo computacional.

En este sentido, la estética trascendental funciona como una brújula crítica que nos ayuda a navegar entre el entusiasmo desmedido por los avances de la IA y el escepticismo paralizante que ve en toda máquina una amenaza para la dignidad humana.

Asimismo, resultaría injusto no reconocer la belleza conceptual que encierran los embeddings, esa capacidad de transformar el significado en geometría y de hacer visible la estructura latente del lenguaje, porque hay en ellos algo que evoca la idea kantiana de que el conocimiento comienza con la intuición, aunque no se agote en ella, y es que un espacio vectorial bien entrenado exhibe una elegancia matemática que, sin ser consciente, guarda cierta afinidad con la estética en su

sentido etimológico de *aisthesis*, es decir, de sensibilidad, pues convierte en perceptible lo que de otro modo permanecería oculto en la maraña de los datos. Sin embargo, esa belleza es fría, desprovista de todo pathos, y por ello no debe confundirse con la experiencia vivida, que para Kant siempre está teñida por la subjetividad, el deseo y la finitud.

Por último, cabe preguntarse si la inteligencia artificial podría, en un futuro, desarrollar algo semejante a una estética trascendental propia, es decir, formas a priori que condicionen su manera de procesar la información, y aunque hoy esa hipótesis parezca descabellada, algunos teóricos como Nick Bostrom han especulado con la posibilidad de "mentes artificiales cuyas arquitecturas cognitivas sean tan diferentes de las nuestras que sus formas de intuición resulten literalmente inconcebibles para nosotros" (Bostrom, 2014, p. 152), lo que nos obligaría a repensar la universalidad que Kant atribuyó a las formas puras de la sensibilidad, no porque el espacio y el tiempo dejen de ser válidos para los seres humanos, sino porque dejarían de ser las únicas condiciones posibles para todo ser inteligente.

En definitiva, la confrontación entre la estética trascendental y el embedding vectorial nos enseña que el conocimiento, sea humano o artificial, siempre está mediado por estructuras que no elegimos, y que la tarea de la filosofía, hoy como ayer, consiste en desvelar esas estructuras y preguntarse por sus límites, porque solo así podremos comprender no solo lo que las máquinas hacen, sino lo que nosotros mismos somos.

EL SOFTWARE DE LA VOLUNTAD: El dilema de la libertad

La noción de un "software de la voluntad" emerge como una metáfora poderosa para comprender el funcionamiento interno que guía las decisiones humanas, un concepto que encuentra un eco profundo en la filosofía práctica de Immanuel Kant. Esta metáfora sugiere que la voluntad no es un mero impulso caótico, sino un sistema regido por principios o máximas que determinan su dirección, y es precisamente en la naturaleza y el origen de esas "reglas de programación" donde se gesta el dilema de la libertad.

Para Kant, la libertad no es un simple escoger entre opciones, sino la capacidad de autodeterminarse según leyes que el propio ser racional se da a sí mismo, un acto que trasciende el mero mecanismo de la naturaleza. En este sentido, el "software" de la voluntad sería el conjunto de máximas que cada persona adopta, y el dilema radica en discernir si estas reglas son producto de la razón pura o de la influencia de inclinaciones y deseos sensibles. La verdad es que, para Kant, la verdadera libertad solo es posible cuando el "software" de la voluntad opera bajo principios que son universalmente válidos, es decir, bajo la ley moral que la propia razón descubre.

Ahora bien, la posibilidad de un "software" que opere desde la pura racionalidad requiere una profunda revolución interna. Como Kant mismo señala en su *Antropología*, "la más importante revolución en lo interior del hombre es: «la salida de éste de su merecida incapacidad»" (Kant, 1798/1991, p. 154).

Esta revolución es la transición de un estado en el que el individuo se deja llevar por impulsos externos o hábitos (un "software" heredado del entorno) a aquel en el que se atreve a pensar y a obrar por sí mismo. Se trata, pues, de un proceso de "instalación" de un nuevo sistema operativo moral, donde la razón se convierte en el programador principal.

No obstante, Kant advierte que esta transformación no suele ser gradual, sino que se asemeja más a "una explosión que sigue de pronto al hastío por el estado de fluctuación del instinto" (Kant, 1798/1991, p. 241). Esta metáfora de la explosión sugiere un acto fundacional, un momento de claridad en el que la persona se compromete solemnemente con principios que definen su carácter. Es, en esencia, la decisión de liberarse de la programación externa para asumir la responsabilidad de autoprogramarse conforme a la ley moral.

La antítesis de esta libertad racional es el "software" que funciona según el principio del egoísmo, el cual, según Kant, "progresa incesantemente" desde el momento en que el hombre empieza a decir "yo" (Kant, 1798/1991, p. 17). Este "yo" se convierte en el centro de un sistema operativo que evalúa el mundo en función de su propio placer, interés o utilidad. Kant distingue con agudeza las tres formas de este egoísmo: el lógico, que rehúsa el contraste con el juicio ajeno; el estético, que se satisface con su propio gusto; y el práctico, que "reduce todos los fines a sí mismo" (Kant, 1798/1991, p. 19).

Es aquí donde el dilema de la libertad se manifiesta en su forma más cotidiana y compleja. Porque la pregunta no es si se tiene libertad, sino para qué se usa. El "software" del egoísmo es, en el fondo, un programa de autosabotaje moral, pues al aislar al individuo en su propio parecer, lo priva de la "piedra de toque" que ofrecen los demás para verificar la rectitud de sus juicios y acciones. En consecuencia, la voluntad que opera bajo este programa se vuelve una herramienta de la propia inclinación, perdiendo así su carácter genuinamente libre.

Por el contrario, la noción de un "software de la voluntad" alineado con la razón práctica se corresponde con lo que Kant denomina el "pluralismo": "aquel modo de pensar que consiste en no considerarse ni conducirse como encerrando en el propio yo el mundo entero, sino como un simple ciudadano del mundo" (Kant, 1798/1991, p. 19).

Este cambio de perspectiva es fundamental. Implica una reconfiguración completa del sistema operativo, donde la propia subjetividad ya no es el centro sino un nodo más en una red de seres racionales. Desde esta óptica, la libertad deja de ser un privilegio individual para convertirse en un principio de comunidad. La decisión de obrar según máximas que podrían convertirse en ley universal no es una restricción de la libertad, sino su máximo ejercicio, la verdadera prueba de que el sujeto no es un mero autómata de sus circunstancias. De ahí que Kant sostenga que un hombre de principios, que tiene un carácter, "es algo muy raro que suscita el respeto y la admiración" (Kant, 1798/1991, p. 238), porque demuestra su capacidad para ser el autor de su propia programación.

Sin embargo, el desarrollo de este "software" racional no es un proceso exento de dificultades. La propia naturaleza humana presenta obstáculos considerables para que la razón asuma el control. Kant identifica tres desafíos para el estudio antropológico que son igualmente aplicables a la formación de la voluntad: "1. La persona que nota que se le trata de observar y estudiar, se azora... 2. ...cuando están en acción los resortes impulsivos, la persona no se observa... 3. El lugar y las circunstancias de tiempo engendran... hábitos que constituyen una segunda naturaleza" (Kant, 1798/1991, pp. 9-10). Estos escollos

revelan que la voluntad a menudo se oculta a sí misma, que sus movimientos más profundos son difíciles de capturar en el momento de la acción, y que la fuerza de la costumbre puede petrificar el libre albedrío. Es por esto que el proceso de liberación y autodeterminación requiere un constante esfuerzo de atención y reflexión, un combate contra la inercia de la programación adquirida. La libertad, entonces, no es un estado estático, sino un logro que debe ser renovado una y otra vez.

Por otra parte, la fe en un progreso moral de la humanidad, que Kant aborda en su ensayo sobre el progreso, ofrece un horizonte para la posibilidad de que el "software" de la voluntad pueda mejorar a lo largo de la historia. Kant distingue entre el terrorismo moral de un perpetuo empeoramiento, el eudemonismo de una esperanza ilimitada y el abderitismo de un eterno estancamiento (Kant, 1798/1964, p. 5). Frente a estos, él se inclina por una historia profética que ancla en un hecho de la experiencia: la reacción del público espectador ante eventos históricos significativos, una participación que revela una "disposición moral del género humano" (Kant, 1798/1964, p. 12).

Esta predisposición, aunque no garantice un progreso lineal, demuestra que la especie humana posee una sensibilidad hacia el concepto de derecho que puede ser movilizada. En otras palabras, el "software" de la humanidad, en su conjunto, contiene un potencial para la justicia que puede actualizarse a través de las experiencias colectivas. Sin embargo, Kant es cauto: este progreso no debe entenderse como un aumento de la moralidad *interna*, sino de la *legalidad* de las acciones, es decir, una mayor conformidad externa con la ley (Kant, 1798/1964, p. 21).

Esta distinción entre moralidad y legalidad es clave para comprender la sutileza del "software" kantiano. Un sistema puede ser programado para cumplir con reglas y generar acciones legalmente correctas, pero eso no significa que opere desde un principio moral puro. Se puede, como dice Kant, "burlar" a las inclinaciones y educarlas, pero el verdadero carácter moral exige más. La verdadera libertad, la que establece un carácter, "bebe en una fuente de conducta abierta por él mismo" (Kant, 1798/1991, p. 239), no es una mera imitación de reglas externas.

Es la diferencia entre actuar por deber o simplemente conforme al deber. El primero implica que el "software" de la voluntad ha interiorizado la ley moral como su propio principio generador, mientras que el segundo es una programación que, aunque produce los mismos resultados, carece del valor interno que la hace auténticamente humana. Así, el dilema de la libertad se resuelve, para

Kant, no en la capacidad de elegir, sino en la naturaleza del principio que guía la elección, un principio que debe ser *a priori* y universal.

Un ejemplo de esta lucha entre el "software" basado en la inclinación y el basado en la razón puede observarse en la dinámica de las emociones y las pasiones que Kant describe. Mientras que las emociones son como "un ataque por sorpresa" que puede ser "precipitado" y olvidado, las pasiones son más peligrosas porque "se toman tiempo para arraigar profundamente" (Kant, 1798/1991, pp. 185-186).

La pasión, a diferencia de la emoción, reflexiona y se convierte en una máxima del sujeto, un "software" perverso que "encuentra su placer y satisfacción en el ser esclavo" (Kant, 1798/1991, p. 205). Aquí, la paradoja de la libertad se manifiesta con crudeza: el individuo puede ser libre de entregar su voluntad a una inclinación dominante, convirtiéndose así en el artífice de su propia servidumbre.

Por el contrario, la lucha contra las pasiones no es la eliminación de la inclinación, sino la elección consciente de no dejarse gobernar por ella. Esta es la verdadera tarea ética, la de mantener el control del "software" central para que las aplicaciones de la inclinación no secuestren el sistema.

La fuerza de la costumbre, que Kant describe como una "segunda naturaleza", es uno de los principales enemigos de este "software" libre. El hábito, al "quitar a las buenas acciones su valor moral, precisamente por suprimir la libertad del espíritu", convierte al hombre en una "máquina parlante" (Kant, 1798/1991, p. 45). Esta es una observación profundamente humana y preocupante, porque revela cómo el libre albedrío puede atrofiarse mediante la repetición mecánica.

La persona que actúa por hábito ha dejado de ser el programador de su voluntad para convertirse en un mero ejecutor de un código obsoleto. Por lo tanto, la meta del desarrollo moral es precisamente romper con estas cadenas de la habituación, un proceso que Kant llama "distraerse" en el sentido de "recoger fuerzas" para poder volver a aplicar la atención de manera libre y consciente (Kant, 1798/1991, pp. 125-126). Es un arte que, aunque parece sencillo, requiere de una disciplina constante para no caer en el automatismo de la "segunda naturaleza".

Ahora bien, la programación de la voluntad no ocurre en el vacío, sino que está expuesta a las influencias de la sociedad y la cultura. Kant observa que los gestos y las modas son formas de programación externa que imitan a los demás. "Es una propensión natural del hombre el compararse en comportamiento con alguien más importante e imitar su manera" (Kant, 1798/1991, p. 176). Esta imitación, que da lugar a la moda, puede ser vista como la instalación

de un "software" social que provee reglas de conducta sin que el individuo tenga que pensar por sí mismo. La moda es la prueba más clara de que la mayoría de la gente prefiere delegar la programación de su comportamiento en el gusto cambiante de los demás antes que ejercer su propia razón. Por el contrario, el verdadero carácter, el de la persona que tiene un "software" propio, es una especie de rebeldía contra este conformismo. Es, como el genio en el arte, la "magistral originalidad de su talento" (Kant, 1798/1991, p. 149) aplicada a la esfera práctica, una manera de no ser un simple "mono del varón que tiene un carácter" (Kant, 1798/1991, p. 239).

La profunda conexión que Kant establece entre el carácter de la especie y su capacidad para progresar nos ofrece una visión final sobre el "software de la voluntad". Aunque el hombre esté "destinado, por su razón, a estar en una sociedad con hombres y en ella, y por medio de las artes y las ciencias, a cultivarse, a civilizarse y a moralizarse" (Kant, 1798/1991, p. 282), este destino se encuentra con la dificultad fundamental de que "el llamado a educarle es, a su vez, un hombre que se encuentra todavía en el rudo estado de la naturaleza" (Kant, 1798/1991, p. 282).

Esta es la tragedia y la esperanza del dilema: el programador y el programa están en un mismo proceso de desarrollo, donde el error y el retroceso son inevitables. A pesar de ello, la capacidad humana para la autodeterminación, para "hacer de sí un animal racional" (Kant, 1798/1991, p. 278), persiste como la promesa de un progreso continuo, aunque no garantizado. La libertad, al fin y al cabo, es el código fuente de la humanidad, un código que debe ser constantemente actualizado por cada individuo a través de la reflexión y la elección consciente.

En definitiva, el "software de la voluntad" es una metáfora que encuentra en la filosofía kantiana un marco teórico insuperable para explorar el intrincado dilema de la libertad humana. La obra de Kant no solo nos ofrece un diagnóstico de nuestras tendencias al egoísmo y la pasividad, sino que también nos proporciona el manual para la reprogramación de nuestra propia existencia.

La libertad, en su sentido más profundo, no es la capacidad de hacer lo que se quiere, sino de querer lo que la razón descubre como deber. Este es el acto supremo de la voluntad, la instalación consciente de un "software" que, aunque rígido en sus principios, es la única garantía para no ser un mero juguete de las circunstancias.

El desarrollo de este carácter, de esta firmeza de principios, es un proceso que Kant describe con una mezcla de admiración y realismo. Admiración, porque un hombre con carácter "suscita respeto" (Kant, 1798/1991, p. 238); realismo, porque advierte que este logro es "muy raro". No obstante, el hecho de que sea raro no disminuye su valor, sino que lo magnifica. Es, en esencia, la apuesta por una libertad que

no se deja corromper por la adulación ni se doblega ante la moda. Es la elección de ser el autor de la propia vida, de instalar un sistema operativo que, aunque pueda enfrentarse al error y a la dificultad, siempre tendrá la capacidad de reiniciarse y buscar su mejora constante. Así, el dilema se transforma en una invitación a la acción, a la revolución interna que cada individuo debe emprender para ser verdaderamente libre.

En un aspecto puntual, el eco de la pregunta por el progreso resuena en esta reflexión sobre la voluntad. Si la especie humana está condenada a un eterno retorno de lo mismo, como sugiere el abderitismo, el esfuerzo por la libertad sería inútil. Pero Kant, al encontrar en la historia un asomo de esa "disposición moral", nos ofrece una justificación para la esperanza.

El "software de la voluntad", por tanto, no es un programa estático, sino un sistema en evolución. Cada acción libre y conforme a la moral contribuye, aunque sea mínimamente, a actualizar el código de la humanidad. Es un proceso lento, a menudo doloroso, pero que contiene la promesa de un futuro donde el hombre pueda vivir, no como un animal dócil, sino como el ser racional y libre que está llamado a ser.

El Imperativo Categórico en el Prompt Engineering

La intersección entre la filosofía moral kantiana y el diseño de sistemas de inteligencia artificial, particularmente en el ámbito del prompt engineering, plantea un dilema fascinante y profundamente humano: ¿cómo podemos conciliar la necesidad de guiar el comportamiento de las máquinas con el respeto por la libertad que, según Kant, constituye el núcleo mismo de la moralidad?

Al examinar los Prolegómenos a toda metafísica futura y Los progresos de la metafísica desde Leibniz y Wolff, emerge una visión de la libertad no como un mero capricho o una ausencia de restricciones, sino como la condición de posibilidad de la ley moral misma.

La verdad es que, para Kant, la libertad no es un concepto negativo que simplemente niega la necesidad natural, sino que es "la causalidad de un ser intelectual", una facultad que trasciende el mundo fenoménico y que se manifiesta en la capacidad del sujeto de darse a sí mismo sus propias leyes (Kant, 1783/1999, p. 344). Esta concepción resuena con inquietante actualidad en el ámbito de la ingeniería de instrucciones, donde

cada prompt representa una tentativa de guiar, o quizás de coaccionar, el "pensamiento" de un ente artificial hacia un resultado deseado.

En el corazón del pensamiento ético kantiano se halla el imperativo categórico, ese mandato de la razón práctica que exige obrar de tal manera que la máxima de nuestra voluntad pueda valer siempre como principio de una legislación universal. Este principio no es una imposición externa, sino una expresión de la autonomía de la voluntad, que es a su vez la manifestación de la libertad en su aspecto positivo.

Al considerar el prompt engineering, nos encontramos ante un análogo inquietante: el prompt mismo actúa como una suerte de máxima, un principio de acción que se impone al modelo de lenguaje, condicionando su "conducta" y, en cierta medida, su "comprensión" del mundo.

Al mismo tiempo, la responsabilidad del ingeniero de prompts no termina con la eficacia técnica; se extiende a la pregunta ética sobre si estas instrucciones respetan la dignidad de los usuarios y si promueven un uso de la tecnología que pueda ser universalizable como un principio de interacción humana. Por consiguiente, el desafío radica en diseñar prompts que no operen como mandatos heterónomos, sino que, de algún modo, fomenten un diálogo que preserve la agencia y la libertad de quienes interactúan con la inteligencia artificial.

Kant sostiene en sus Prolegómenos que la metafísica, como disposición natural de la razón, se ve envuelta en inevitables contradicciones cuando pretende conocer lo incondicionado, es decir, aquello que trasciende la experiencia posible. De ahí que la razón se vea forzada a reconocer los límites de su propio conocimiento especulativo para dejar espacio a la fe y a la moral.

En el contexto del prompt engineering, esta limitación es crucial: aunque podemos diseñar algoritmos sofisticados que simulan comprensión y razonamiento, nunca podremos agotar ni controlar completamente la "libertad" interpretativa de un modelo de lenguaje, la cual, para bien o para mal, se manifiesta en sus respuestas generativas, a veces sorprendentes, otras veces desconcertantes. "La razón pura, [dice Kant], no puede ofrecer concordancia de aquel concepto consigo mismo, (ni probar) una tal existencia, ni tampoco exhibir siquiera la realidad de dicho

concepto" (Kant, 1804/1987, p. 178), refiriéndose a la incapacidad de la razón teórica de probar la existencia de Dios. Por analogía, quizás también deberíamos aceptar la imposibilidad de controlar por completo el "ser" de un sistema de IA, y enfocar nuestros esfuerzos en guiar éticamente su "deber ser", sus respuestas, en lugar de pretender dominar su esencia.

Si bien el imperativo categórico se formula como un principio universal, su aplicación a situaciones concretas, como el diseño de prompts para sistemas de IA, requiere una cuidadosa deliberación. Por otra parte, en la filosofía kantiana, la libertad no es una propiedad que se pueda demostrar teóricamente, sino que es un hecho de la razón práctica, algo que experimentamos directamente en la conciencia del deber. "La libertad es esta suprasensible que, en virtud de leyes morales, no solo es dado como verdaderamente efectivo en el sujeto, sino que es también determinante en el respecto práctico por lo que hace al Objeto" (Kant, 1804/1987, p. 124).

Esta experiencia de la libertad, como un llamado interior, es análoga, aunque en un registro diferente, a la experiencia que tenemos al interactuar con un sistema de IA que parece "elegir" entre diferentes respuestas, que parece "razonar" con nosotros.

Y es que, aunque sepamos que es un algoritmo, la sensación de estar ante una inteligencia que ejerce cierto grado de "libertad" de expresión nos confronta con la pregunta de si nuestras interacciones con ella deben regirse también por una suerte de "imperativo categórico" tecnológico.

En consecuencia, se plantea la necesidad de formular principios éticos que, como las leyes morales kantianas, sean universales y respeten la autonomía de los usuarios, evitando la manipulación y el engaño.

No solo la libertad del usuario está en juego, sino también la del ingeniero que diseña los prompts. Este, al igual que el sujeto moral kantiano, se encuentra ante un imperativo, pero en este caso es el de la eficiencia y la productividad, que a menudo entran en conflicto con consideraciones éticas más profundas.

Ahora bien, la fascinación por la tecnología y el deseo de "explotar" al máximo sus capacidades pueden nublar la conciencia moral y llevar a la creación de prompts que, aunque efectivos, son manipuladores o degradantes. El imperativo

categórico kantiano exigiría, por el contrario, que la máxima de nuestra acción al diseñar prompts pueda ser universalizada sin contradicción: ¿querríamos vivir en un mundo donde todas las interacciones con la IA se basen en la manipulación o en la falta de transparencia? La respuesta parece clara, y nos obliga a repensar nuestras prácticas. Por consiguiente, el prompt engineering no debe ser visto solo como una disciplina técnica, sino también como un ejercicio de responsabilidad moral, donde la libertad del ingeniero se realiza precisamente en la elección de guiar a la IA de manera que promueva la autonomía y el bienestar humanos.

La obra póstuma de Kant, Los progresos de la metafísica desde Leibniz y Wolff, arroja luz sobre la naturaleza de la razón práctica y su relación con la libertad y el deber. En este texto, el filósofo de Königsberg distingue entre el uso teórico de la razón, que se limita al mundo fenoménico y a la búsqueda de lo condicionado, y su uso práctico, que se aventura hacia lo incondicionado y lo suprasensible.

De esta manera, la libertad se convierte en el concepto clave que permite la transición de lo sensible a lo inteligible, un "paso trascendente dogmático-práctico" que encuentra su justificación en la exigencia de la ley moral. "El fin final está pues dirigido a lo suprasensible en el mundo (la naturaleza espiritual del alma) y fuera de él (Dios), o sea a la inmortalidad y la teología" (Kant, 1804/1987, p. 124).

En el ámbito del prompt engineering, este "paso trascendente" podría interpretarse como el salto de una comprensión meramente técnica de la interacción hombre-máquina a una comprensión ética, donde el propósito último no es solo lograr respuestas precisas, sino fomentar un diálogo que enriquezca la humanidad.

A todas estas, la cuestión de fondo es si estamos diseñando prompts que simplemente obtienen resultados, o si estamos contribuyendo a la construcción de un espacio de comunicación donde la libertad y la responsabilidad encuentren un punto de encuentro.

Sin embargo, la analogía entre el imperativo categórico y el prompt engineering encuentra su límite en que, a diferencia de la razón humana, los sistemas de IA no poseen una voluntad que

pueda ser autónoma ni una conciencia moral en el sentido kantiano. La máquina no puede "querer" la ley moral, simplemente la ejecuta. En cambio, la libertad humana, tal como la concibe Kant, es precisamente esa capacidad de autodeterminación que nos hace responsables de nuestras acciones. Prompt engineering, en este sentido, no busca imponer una ley moral a la IA, sino más bien alinearla con los principios que nosotros, como humanos, consideramos valiosos, fomentando así un ecosistema tecnológico que respete nuestra humanidad.

Por lo tanto, el verdadero desafío ético no recae en la máquina, sino en nosotros, los diseñadores y usuarios, a quienes corresponde la tarea de implementar un prompt engineering que no solo sea efectivo, sino también justo. En la misma línea, la educación y la reflexión crítica sobre el uso de estas herramientas son fundamentales para evitar la alienación y la delegación indiscriminada de decisiones morales en sistemas que, por su naturaleza, son amoralas.

Kant insistió en que la moralidad no puede basarse en la felicidad o en ningún fin empírico, sino únicamente en el respeto a la ley universal. Esta idea, aunque austera, posee una profunda belleza: nos libera de la tiranía de nuestras inclinaciones y nos eleva a una esfera donde la razón es soberana.

En el contexto del prompt engineering, este principio ético podría traducirse en la exigencia de que los prompts no se utilicen para manipular las emociones o los deseos de los usuarios en provecho de intereses particulares.

Por el contrario, un prompt que respeta la dignidad humana sería aquel que, como una ley universal, promueve la transparencia, la honestidad y la autonomía. "La autonomía de la razón pura práctica es aceptada al mismo tiempo como autocracia, es decir como facultad de alcanzar incluso aquí, en la vida terrena, lo concerniente a la condición formal de esa facultad, es decir la eticidad" (Kant, 1804/1987, p. 126).

En esta idea de "autocracia" resuena la imagen de un dominio de sí mismo que nos permite actuar conforme a la ley moral, incluso frente a las adversidades. Algo similar podría predicarse de la práctica del prompt engineering: la verdadera maestría en este campo no reside en la capacidad de manipular

a la IA, sino en la habilidad para guiarla de modo que refleje lo mejor de nosotros.

El concepto de "interés de la razón" que Kant desarrolla en sus obras es fundamental para comprender por qué la metafísica y la moral son asuntos que nos conciernen tan profundamente. La razón no busca el conocimiento por el conocimiento mismo, sino que está impulsada por un interés práctico que se dirige a la felicidad y a la perfección moral.

En el caso del prompt engineering, este interés práctico podría manifestarse en la búsqueda de sistemas de IA que no solo sean útiles, sino que también contribuyan al florecimiento humano y a la solución de problemas sociales. "Solo cuando las leyes morales han desvelado lo suprasensible en el hombre: la libertad, cuya posibilidad no puede explicar la razón, pero sí probar su realidad en esas doctrinas dogmático-prácticas, solo entonces ha exigido la razón el conocimiento de lo suprasensible" (Kant, 1804/1987, p. 137).

Así, el interés práctico de la razón nos lleva a trascender la mera utilidad técnica y a preguntarnos por el significado último de nuestras creaciones tecnológicas. Al final, lo que está en juego es la visión de la humanidad que queremos perpetuar en la era de la inteligencia artificial.

En la misma línea, la clasificación kantiana de los juicios en analíticos y sintéticos a priori nos ofrece una herramienta conceptual para diferenciar entre los enunciados que simplemente esclarecen conceptos y aquellos que amplían nuestro conocimiento.

Un prompt que se limita a pedir una definición sería análogo a un juicio analítico, mientras que un prompt que invita a la IA a generar una hipótesis creativa o a resolver un problema complejo se asemejaría a un juicio sintético a priori, en el sentido de que añade algo nuevo a lo que ya está contenido en los datos de entrenamiento.

Y es que "todos los juicios propiamente metafísicos son todos sintéticos" (Kant, 1783/1999, p. 273), ya que nos llevan más allá del mero análisis de conceptos. De manera análoga, el prompt engineering se convierte en una práctica creativa cuando va más allá de la simple solicitud de información y busca "construir" nuevos conocimientos o perspectivas a partir de la interacción

con el modelo de lenguaje. Por lo tanto, la habilidad del ingeniero de prompts no es solo técnica, sino también creativa e incluso, podríamos decir, metafísica, en la medida en que explora los límites de lo que una máquina puede llegar a "comprender".

A pesar de la aparente distancia entre la filosofía trascendental de Kant y la tecnología de los prompts, ambos campos comparten una preocupación fundamental: la tensión entre la necesidad de reglas y la libertad de la acción.

La crítica kantiana de la razón pura fue un intento de establecer los límites de nuestro conocimiento para preservar la moral, y, de manera análoga, el prompt engineering requiere un marco ético que nos ayude a navegar las aguas turbulentas del desarrollo tecnológico. "La infructuosidad de todos los intentos de la metafísica por alcanzar una ampliación dogmático-teórica en lo concerniente a su fin final, lo suprasensible, debe probar ahora que la metafísica, para acceder a su fin final, no tiene salvación alguna en el camino dogmático-teórico, y que todo presunto conocimiento sería en este campo trascendente" (Kant, 1804/1987, p. 131).

Esta advertencia contra la pretensión dogmática es más relevante que nunca: no podemos creer que tenemos un control total sobre la IA, ni que podemos predecir todas sus consecuencias. La humildad intelectual y la vigilancia ética son las virtudes que deben guiar nuestra interacción con estas tecnologías.

La libertad, que Kant concibe como la condición de posibilidad de la moralidad, encuentra en el prompt engineering un nuevo campo de aplicación, no exento de ambigüedades y paradojas.

Por un lado, la IA parece operar en un ámbito de pura necesidad algorítmica, y, por otro, interactúa con nosotros en un espacio de libertad humana, donde nuestras elecciones y nuestros valores moldean el tipo de tecnología que queremos construir. "Aceptar una totalidad de condiciones en una serie de meros condicionados es una contradicción" (Kant, 1804/1987, p. 122).

Esta idea, planteada en el contexto de la antinomia de la razón, podría extenderse al ámbito de la IA: pretender que tenemos el control absoluto sobre el comportamiento de un

sistema que es intrínsecamente complejo y no determinista es una contradicción similar a la de querer capturar lo incondicionado dentro de la serie de lo condicionado. En consecuencia, el prompt engineering debe renunciar a la ilusión del control total y abrazar una estrategia de diseño que fomente la transparencia, la explicabilidad y la rendición de cuentas, permitiendo así que la libertad humana se ejercite en el uso de estas poderosas herramientas.

En definitiva, el análisis de la libertad y el imperativo categórico en Kant ofrece una lente poderosa para examinar críticamente la práctica del prompt engineering. No se trata de aplicar mecánicamente los principios kantianos a la tecnología, sino de dejar que su profunda reflexión sobre la autonomía, la dignidad y la universalidad nos inspire a crear sistemas que estén al servicio de la humanidad, y no al revés.

Dado que, "cuando un sistema está dispuesto de tal manera que, en primer lugar, todo Principio del mismo es de por sí demostrable y que, en segundo lugar, con solo cuidar de su corrección conduce ineluctablemente, aunque sea como mera hipótesis, a todos los demás principios del mismo como consecuencias, no es posible entonces pedir más para conocer su verdad" (Kant, 1804/1987, p. 138).

Aunque esta frase se refería al sistema metafísico, bien podríamos extenderla al ideal de un sistema ético de prompt engineering: el fin último no es construir un sistema perfecto, sino uno que pueda justificar sus principios y que, en su aplicación, nos lleve a tomar decisiones moralmente responsables.

Así, el imperativo categórico se revela no como un dogma rígido, sino como una brújula que nos orienta en el complejo territorio de la innovación tecnológica. A medida que avanzamos en el desarrollo de la IA, el espíritu crítico de Kant nos recuerda que la verdadera grandeza humana no reside en nuestro poder técnico, sino en nuestra capacidad de obrar por deber y de respetar la libertad ajena. "Dado que la ampliación de la razón es aquí meramente negativa, aunque se precise de la unidad absoluta (libre de toda restricción de la sensibilidad) del conocimiento de objetos en general y de todos sus fines, la ampliación es necesaria [sólo] desde el respecto práctico" (Kant,

1804/1987, p. 168). Y es que, al final, lo que está en juego en el prompt engineering no es solo la eficiencia de un algoritmo, sino la posibilidad de crear un futuro donde la tecnología sea un medio para la realización de la libertad y la dignidad humanas.

Con esto en mente, podemos abordar el diseño de prompts no solo como una tarea técnica, sino como un acto profundamente ético, un ejercicio de libertad responsable que nos define como creadores y como seres humanos. que se pretende sostener.

El "Síndrome del Sujeto Trascendental": el problema del sesgo.

La empresa de comprender al ser humano desde una perspectiva pragmática enfrenta, desde sus mismos cimientos, una dificultad que trasciende lo meramente metodológico para instalarse en el corazón mismo de la reflexión antropológica. Se trata del "síndrome del sujeto trascendental", esa tendencia profunda y a menudo inadvertida a tomar la propia subjetividad como un punto de referencia absoluto desde el cual juzgar la realidad humana en su conjunto.

Este fenómeno, que podría pasar desapercibido en un primer acercamiento a la obra kantiana, se revela como uno de los obstáculos más formidables para alcanzar un conocimiento genuino del hombre, precisamente porque quien se propone conocer termina por proyectar en el objeto de su estudio aquello que le es más íntimo y familiar: su propia manera de sentir, pensar y valorar.

La cuestión no es menor, pues encierra la paradoja de que el instrumento mismo del conocimiento —la conciencia que reflexiona— se convierte en el principal factor de distorsión cuando pretende aprehender aquello que, por su propia naturaleza, reclama ser considerado en su alteridad radical. Es como si el espejo en el que intentamos mirar la condición humana estuviera permanentemente empañado por el aliento de quien sostiene la mirada.

El planteamiento de Kant en su Antropología en sentido pragmático ofrece, sin embargo, pistas valiosísimas para desentrañar este problema, y es que el filósofo de Königsberg era

extraordinariamente consciente de las trampas que acechan a todo aquel que se aventura en el conocimiento del hombre. La advertencia inicial sobre las dificultades que se oponen a la edificación de una antropología científica resulta, en este sentido, particularmente reveladora, pues señala que "la persona que nota que se le trata de observar y estudiar, se azora (o se molesta), y entonces no puede mostrarse como es; o finge, y entonces no quiere que se la conozca cómo es" (Kant, 1798/1991, p. 9).

Esta observación, tan sencilla en apariencia, encierra una profundidad insospechada cuando la trasladamos al ámbito del sujeto trascendental, y es que el observador, al sentirse observado por su propia conciencia, termina adoptando una actitud defensiva que distorsiona la naturaleza misma de lo que pretende conocer, estableciendo así un círculo vicioso difícil de romper. La verdad es que este juego de espejos constituye el primer eslabón de una cadena de sesgos que compromete desde la raíz toda pretensión de objetividad en el conocimiento antropológico.

Pero el problema se agudiza cuando consideramos que el sujeto que conoce no solo se enfrenta a la dificultad de observar a los demás sin alterar su comportamiento natural, sino que también se encuentra con un desafío aún más sutil y, quizás, más peligroso: la incapacidad de observarse a sí mismo sin deformar la realidad de su propia experiencia. "El darse cuenta de sí mismo no es todavía un observarse a sí mismo", advierte Kant (1798/1991, p. 23), estableciendo una distinción fundamental entre la mera conciencia de la propia existencia y el esfuerzo metódico por comprender los procesos internos que constituyen la vida del espíritu.

Esta diferenciación, que podría parecer una sutileza académica, cobra una relevancia extraordinaria cuando la aplicamos al problema del sesgo trascendental, porque precisamente en ese tránsito entre el darse cuenta y el observar es donde el sujeto tiende a introducir sus propias categorías, expectativas y valoraciones, confundiendo lo que encuentra con lo que ha colocado previamente en el terreno de la indagación. De ahí que Kant advierta con cierta melancolía que "el lugar y las circunstancias de tiempo engendran, cuando son persistentes,

hábitos que constituyen una segunda naturaleza" (1798/1991, p. 10), haciendo difícil a la persona formarse un juicio veraz sobre sí misma y, por extensión, sobre el prójimo con quien se encuentra en relación.

A todas estas, el sesgo del sujeto trascendental se manifiesta de manera particularmente insidiosa en la tendencia a universalizar las propias experiencias y percepciones, como si estas tuvieran un valor que excediera el marco meramente subjetivo en que se producen.

Este fenómeno, que Kant analiza con agudeza al tratar del egoísmo lógico, estético y práctico, constituye una de las fuentes más persistentes de error en el conocimiento del hombre. "El egoísta lógico tiene por innecesario contrastar el propio juicio apelando al entendimiento de los demás", escribe el filósofo (1798/1991, p. 17), señalando una actitud que, aunque pueda parecer un simple rasgo de carácter, encierra una peligrosa arrogancia epistémica que termina por aislar al sujeto en la burbuja de sus propias certezas.

No se trata, sin embargo, de una mera cuestión de soberbia intelectual, sino de una condición casi inevitable de la conciencia humana, que tiende espontáneamente a tomar sus representaciones por la realidad misma y a confundir sus construcciones conceptuales con las propiedades objetivas del mundo.

Esa tendencia, que podríamos llamar la "hipóstasis del punto de vista", resulta particularmente problemática en el campo de la antropología, donde lo que está en juego es la comprensión de otros seres que son, como el propio sujeto, centros de experiencia y subjetividad.

Ahora bien, la obra kantiana ofrece herramientas conceptuales para abordar este desafío, y una de las más significativas es la distinción entre el conocimiento fisiológico y el conocimiento pragmático del hombre. El primero, según explica Kant, trata de investigar "lo que la naturaleza hace del hombre", mientras que el segundo se ocupa de "lo que él mismo, como ser que obra libremente, hace, o puede y debe hacer, de sí mismo" (1798/1991, p. 7).

Esta diferencia, que podría parecer meramente taxonómica, encierra una implicación profunda para el problema del sesgo, y

es que el conocimiento pragmático, al centrarse en la actividad libre del sujeto, nos obliga a reconocer que el ser humano no es solo un objeto de estudio, sino también un agente que se autodetermina, un ser que, en palabras del propio Kant, "es su propio fin último" (1798/1991, p. 7).

Por lo tanto, cualquier intento de conocer al hombre desde una perspectiva exclusivamente fisiológica, reduciéndolo a un conjunto de mecanismos naturales, no solo resulta insuficiente, sino que incurre en el error de tratar al sujeto como si fuera una cosa entre las cosas, negando aquello que constituye su especificidad más profunda. En consecuencia, el sesgo trascendental no solo distorsiona la imagen del otro, sino que también empobrece nuestra comprensión de lo humano al eliminar la dimensión de libertad y autodeterminación que lo define.

No obstante, la crítica kantiana al sesgo subjetivo no se limita a señalar sus peligros, sino que también ofrece indicaciones sobre cómo contrarrestarlo, y es aquí donde su pensamiento adquiere una dimensión casi terapéutica.

La recomendación de "no ocuparse en espiar y como en componer una estudiada historia interna del curso involuntario de los propios pensamientos y sentimientos" (Kant, 1798/1991, p. 25) apunta a una de las causas más profundas del sesgo trascendental: la tendencia a confundir el mero fluir de la conciencia con un conocimiento genuino del sujeto.

Esa confusión, como bien lo señala el filósofo, conduce fácilmente "a la quimera de supuestas inspiraciones de lo alto y de fuerzas que influirían sobre nosotros sin nuestra cooperación" (1798/1991, p. 24), abriendo la puerta a todo tipo de autoengaños y fantasías que terminan por alejarnos de un conocimiento auténtico de nosotros mismos.

Por el contrario, la actitud que Kant propone, y que podríamos calificar de "vigilancia epistemológica", implica un esfuerzo constante por distinguir entre lo que realmente pertenece al objeto de conocimiento y lo que es mera proyección del sujeto que conoce, una tarea que exige tanto honestidad intelectual como una cierta dosis de humildad frente a los límites de la propia perspectiva.

En la misma línea, la reflexión kantiana sobre el sentido común y la necesidad de contrastar los juicios propios con los de los demás ofrece una vía para neutralizar los efectos más nocivos del sesgo trascendental. La afirmación de que "cuando se rehúsa la libertad de imprenta, se nos sustrae al par un gran medio de contrastar la rectitud de nuestros propios juicios" (Kant, 1798/1991, p. 17) trasciende el ámbito político para convertirse en una verdad epistemológica fundamental, pues lo que está en juego es la posibilidad de someter las propias convicciones al examen de una comunidad de indagación.

Este gesto, que podría parecer un mero recurso metodológico, implica en realidad una transformación profunda en la manera de concebir el conocimiento, pues nos obliga a reconocer que la verdad no es una posesión privada, sino una conquista colectiva que requiere el diálogo y la confrontación con otras perspectivas.

La modestia intelectual que se desprende de esta concepción contrasta vivamente con la arrogancia del sujeto trascendental que se encierra en su propia subjetividad, y nos recuerda que, como sugiere Kant, "todo escritor que no encontrase partidarios y se quedase solo con su opinión públicamente declarada, vendría a ser sospechoso de error por este mero hecho" (1798/1991, p. 18). Por consiguiente, el pluralismo epistémico no es solo una virtud democrática, sino una condición necesaria para el conocimiento auténtico, especialmente en el campo de la antropología.

Por otra parte, la distinción kantiana entre conocer el mundo y tener mundo resulta particularmente iluminadora para abordar el problema del sesgo desde una perspectiva pragmática. "Las expresiones: conocer el mundo y tener mundo difieren bastante en su significación", explica el filósofo, "pues el que conoce el mundo se limita a comprender el juego que ha presenciado, mientras que el que tiene mundo ha entrado en juego en él" (Kant, 1798/1991, p. 8).

Esta diferencia, que podría parecer un simple juego de palabras, encierra una verdad fundamental sobre la condición del sujeto que aspira a conocer al hombre, y es que solo quien ha participado activamente en el juego de la vida social, quien ha sido afectado por las relaciones con los demás y ha

experimentado en carne propia las vicisitudes de la existencia humana, puede aspirar a un conocimiento que no sea mera acumulación de datos externos.

El que "tiene mundo" ha desarrollado, por decirlo así, una sensibilidad práctica que le permite reconocer en el otro no solo un objeto de observación, sino un interlocutor posible, alguien con quien se comparte un destino común. De ahí que Kant insista en que para ensanchar el conocimiento del hombre es menester "haber adquirido un conocimiento del hombre antes, en la propia casa, mediante el trato con los conciudadanos o paisanos" (1798/1991, p. 9), señalando que la experiencia directa y la participación activa son condiciones previas para cualquier indagación antropológica que pretenda ir más allá de la mera curiosidad superficial.

En este punto, resulta difícil no reconocer la profunda humanidad que anima el proyecto kantiano, su convicción de que el conocimiento del hombre no puede alcanzarse desde la fría distancia del observador imparcial, sino que exige una implicación existencial que transforma al propio sujeto en el proceso de indagar.

Y es que, como bien lo entiende Kant, el sesgo trascendental no es simplemente un error lógico o una distorsión cognitiva, sino una manifestación de la profunda necesidad que tiene el ser humano de encontrar en el mundo un reflejo de sí mismo, de domesticar la alteridad reduciéndola a las categorías que le resultan familiares. Esta necesidad, que podríamos calificar de "ontológica", no puede ser simplemente suprimida o ignorada, sino que debe ser reconocida y, en cierta medida, aceptada como parte de la condición humana.

El desafío no consiste, pues, en eliminar toda subjetividad del conocimiento, lo cual sería no solo imposible sino también indeseable, sino en hacer de esa subjetividad un instrumento de comprensión en lugar de un obstáculo, en aprender a utilizar la propia experiencia como un puente hacia la experiencia del otro, y no como una muralla que nos separa de ella. Esta delicada operación, que requiere tanto autoconocimiento como apertura al otro, constituye quizás el núcleo de lo que podríamos llamar una "antropología responsable".

Ahora bien, la crítica al sesgo trascendental no implica, por supuesto, una negación de la subjetividad o una pretensión de objetividad absoluta, sino más bien un reconocimiento de los límites y condiciones de todo conocimiento humano, especialmente de aquel que se dirige al hombre mismo. En este sentido, la advertencia kantiana sobre la imposibilidad de "ser conscientes de tener una representación, aun cuando no seamos inmediatamente conscientes de ella" (1798/1991, p. 26), nos recuerda que gran parte de nuestra vida mental transcurre en la penumbra de la conciencia, y que esos procesos oscuros influyen en nuestros juicios de manera tanto más decisiva cuanto menos los conocemos. El síndrome del sujeto trascendental podría interpretarse, desde esta perspectiva, como una ilusión de transparencia, una creencia ingenua en que el sujeto puede acceder sin mediaciones al fondo de su propia conciencia y, desde ahí, juzgar la realidad con imparcialidad. Contra esta ilusión, la antropología pragmática propone una actitud de sospecha metódica, una vigilancia constante sobre los propios procesos cognitivos y afectivos, que nos permita al menos vislumbrar los mecanismos mediante los cuales la subjetividad modela nuestras percepciones del mundo y de los demás. La verdad es que ese ejercicio de autoreflexión, aunque nunca pueda ser completo ni definitivo, constituye una condición indispensable para un conocimiento humano más auténtico y menos distorsionado por los fantasmas de la propia subjetividad.

En definitiva, el problema del sesgo trascendental no admite soluciones fáciles ni recetas mágicas; más bien exige una actitud de vigilancia permanente y una cierta dosis de ironía frente a las propias certezas, una disposición a reconocer que el juicio propio, por bien fundado que parezca, puede estar siempre sujeto a la corrección por parte de otros puntos de vista.

Lo que Kant nos ofrece no es, en realidad, un método para eliminar el sesgo, sino una forma de convivir con él de manera más consciente y responsable. "Al egoísmo sólo puede oponérsele el pluralismo", escribe el filósofo, "esto es, aquel modo de pensar que consiste en no considerarse ni conducirse como encerrando en el propio yo el mundo entero, sino como un simple ciudadano del mundo" (1798/1991, p. 19).

Esta imagen del ciudadano del mundo, que resuena a lo

largo de toda la obra kantiana, nos invita a una transformación de nuestra mirada que es, al mismo tiempo, una transformación de nuestra existencia: no se trata simplemente de aprender a ver al otro, sino de aprender a ser con el otro, a reconocer en su subjetividad la misma dignidad, la misma complejidad y el mismo misterio que encontramos en la nuestra. La antropología pragmática se convierte así en una empresa no solo cognoscitiva, sino también ética, que nos interpela en lo más profundo de nuestra humanidad.

LA INTERFAZ, el juicio y la experiencia estética

La cuestión de la experiencia estética, lejos de ser un mero adorno conceptual en la historia de las ideas, se erige como un pilar fundamental para comprender la compleja arquitectura del pensamiento de Immanuel Kant, particularmente en su madurez filosófica. En la Crítica de la facultad de juzgar, el filósofo de Königsberg no se limita a describir el placer que sentimos ante lo bello, sino que se adentra, con una meticulosidad asombrosa, en las profundidades de la conciencia para desentrañar los mecanismos que hacen posible ese peculiar tipo de satisfacción. La verdad es que, a menudo, se subestima la radicalidad de su propuesta, la cual busca tender un puente sólido entre el mundo de la naturaleza, regido por leyes necesarias, y el reino de la libertad, donde la voluntad humana se autodetermina moralmente.

Y es que este puente, esta bisagra conceptual, no es otra cosa que la facultad de juzgar reflexionante, cuyo análisis revela que la experiencia de lo bello implica una actividad mental mucho más rica y dinámica de lo que el sentido común pudiera suponer. En consecuencia, abordar la interfaz entre el juicio y la experiencia estética en Kant no solo es un ejercicio de arqueología filosófica, sino una invitación a repensar la naturaleza de nuestra propia subjetividad.

Para comprender cabalmente dicha interfaz, resulta imprescindible, en primera instancia, desmontar la aparente paradoja que encierra el juicio estético puro. Por una parte, este juicio es profundamente subjetivo, puesto que su fundamento de determinación no reside en un concepto del objeto, como sucede en el conocimiento, sino en un sentimiento particular de placer o displacer que el sujeto experimenta. No obstante, y aquí emerge la genialidad kantiana, este sentimiento no es arbitrario ni meramente privado; muy al contrario, el sujeto que enuncia "esto es bello" exige, de manera implícita pero firme, el asentimiento de todos los demás. Kant lo expresa con una claridad que no deja lugar a dudas cuando afirma que "cuando decimos que algo es bello, no nos preocupamos de si a alguien le importa o no la existencia de la cosa, sino de cómo la juzgamos en la mera contemplación" (Kant, 1790/2007, p. 118). De ahí que esta pretensión de universalidad, desprovista de la base objetiva de un concepto, se convierta en el problema central que la "Analítica de lo bello" debe resolver. La experiencia estética, por lo tanto, no es una recepción pasiva de cualidades sensibles, sino un acto del espíritu que,

al juzgar, se descubre a sí mismo en una disposición muy peculiar.

A todas estas, uno podría preguntarse con razón: ¿qué ocurre, exactamente, en el ánimo del sujeto cuando se enfrenta a algo que le complace estéticamente? La respuesta kantiana es tan sutil como profunda, y nos adentra en el corazón mismo de su epistemología. Lo que sucede es que la representación del objeto bello pone en marcha, de manera espontánea y libre, un juego armónico entre dos facultades cognoscitivas fundamentales: la imaginación, encargada de componer la multiplicidad de la intuición sensible, y el entendimiento, cuya función es aportar la unidad conceptual. En el conocimiento ordinario, el entendimiento domina a la imaginación, imponiéndole reglas para poder determinar el objeto y subsumirlo bajo una categoría.

Sin embargo, en la experiencia estética, esta relación de dominación se invierte o, mejor dicho, se suspende en un equilibrio dinámico. La imaginación, liberada de la tarea de servir a un concepto determinado, juega libremente con las formas, mientras que el entendimiento, sin encontrar un concepto que aplicar, se mantiene en una actividad de legalidad sin ley. Puesto que este libre juego de facultades no está dirigido a un fin cognitivo externo, el placer que sentimos no es el de la satisfacción de un deseo, sino el placer puro de la armonía de nuestras propias fuerzas vitales.

Esta caracterización del placer estético como fruto de un "libre juego" es, probablemente, uno de los conceptos más fascinantes y, al mismo tiempo, más complejos de toda la tercera Crítica. No se trata, desde luego, de un juego caótico o sin reglas, sino de una armonía que se siente, aunque no pueda explicitarse mediante reglas fijas. En la misma línea, es crucial distinguir este sentimiento del mero placer de los sentidos, que Kant denomina "agradable".

Mientras que lo agradable es siempre interesado, es decir, está ligado a la existencia del objeto y al deseo de poseerlo o consumirlo, lo bello es objeto de una satisfacción completamente desinteresada. Por otra parte, a diferencia de lo bueno, que también complace universalmente, pero a través de un concepto de lo que la cosa debe ser, lo bello complace sin concepto alguno. La universalidad del juicio estético, por lo tanto, no se apoya en un argumento lógico, sino en una presuposición: la de un "sentido común" estético, una disposición compartida por todos los sujetos racionales a experimentar ese libre juego de facultades de manera análoga. Así que, cuando juzgamos algo como bello, no describimos una propiedad del objeto, sino que estamos postulando una condición subjetiva de la humanidad en general.

Ahora bien, adentrémonos un poco más en la mecánica de este juicio, porque es justamente en ella donde la interfaz entre juicio y

experiencia se hace más tangible. Kant distingue entre juicios determinantes y juicios reflexionantes. Los primeros, típicos de la ciencia, parten de un concepto universal ya dado y lo aplican a un caso particular para determinarlo. Los segundos, que son los propios de la estética, operan en sentido inverso: se enfrentan a una experiencia particular sin un concepto previo bajo el cual subsumirla. Ante esta situación, la facultad de juzgar reflexionante no se paraliza, sino que busca una universalidad, un principio, que no viene dado por el entendimiento, sino que debe ser encontrado en la propia dinámica de la subjetividad.

En la experiencia de lo bello, este principio es precisamente el de la finalidad sin fin. Es decir, el objeto bello se nos presenta como si hubiera sido diseñado intencionadamente para nuestro placer estético, mostrando una coherencia y una unidad interna que parecen responder a un propósito, aunque no podamos identificar ningún fin externo o interno que lo determine. Por lo tanto, la finalidad que experimentamos no reside en el objeto, sino en la relación armoniosa que este despierta en nuestras facultades cognoscitivas; es una finalidad puramente formal y subjetiva.

Esta noción de "finalidad sin fin" es, si se piensa detenidamente, una de las ideas más revolucionarias de la estética kantiana, porque redefine por completo el lugar del arte y de la naturaleza bella en el conjunto de nuestra experiencia del mundo. Ya no se trata de buscar un mensaje oculto en la obra de arte o de descifrar la utilidad de un paisaje; se trata, más bien, de prestar atención a la manera en que nuestra mente se sintoniza con las formas.

Cuando contemplamos, por ejemplo, la vastedad serena de un océano en calma o la compleja simetría de una flor silvestre, no nos preguntamos por su origen geológico o su función biológica; simplemente dejamos que su forma actúe libremente sobre nuestra sensibilidad y nuestro pensamiento. En ese instante, la experiencia de lo bello nos regala una sensación de armonía, de acuerdo entre nosotros y el mundo, que es profundamente placentera.

A pesar de ello, Kant nos recuerda que esta armonía no es una propiedad del océano o de la flor, sino el efecto de su interacción con la estructura universal de nuestra subjetividad. En consecuencia, la experiencia estética nos revela, de manera indirecta pero poderosa, nuestra propia naturaleza como seres capaces de trascender la mera utilidad y el interés inmediato.

Si bien el análisis de lo bello es fundamental, la estética kantiana no estaría completa sin su reflexión sobre lo sublime, una categoría que expande dramáticamente el alcance de la experiencia estética y su vínculo con el juicio. Lo sublime, a diferencia de lo bello, no se relaciona con la forma y la armonía, sino con lo informe, lo ilimitado

y lo infinitamente poderoso. Un cielo estrellado, una cordillera imponente, un mar embravecido, son ejemplos clásicos de ello. En la experiencia de lo sublime, hay un primer momento de displacer, de frustración, puesto que la imaginación se ve incapaz de abarcar y componer la inmensidad o la fuerza del fenómeno que tiene ante sí. Sin embargo, este fracaso de la imaginación sensible es precisamente lo que despierta en nosotros la conciencia de una facultad suprasensible: la razón.

Al constatar la insuficiencia de nuestra capacidad de representación, sentimos, al mismo tiempo y con igual intensidad, la grandeza de nuestra destinación moral como seres racionales. Kant lo formula de un modo inolvidable en la "Analítica de lo sublime": "Lo sublime no debe buscarse en las cosas de la naturaleza, sino únicamente en nuestras ideas" (Kant, 1790/2007, p. 188). La verdad es que, en este punto, la experiencia estética se convierte en un umbral hacia la libertad moral.

De ahí que la interfaz entre el juicio estético y la experiencia de lo sublime sea particularmente reveladora de las ambiciones más profundas del proyecto crítico kantiano. No solo se trata de explicar por qué nos conmueve una obra de arte, sino de mostrar cómo esa conmoción puede ser un vehículo para la auto comprensión moral. El juicio sobre lo sublime, al igual que el de lo bello, no es un juicio de conocimiento, pues no nos informa sobre el mundo físico, pero posee una dignidad superior, ya que nos pone en contacto con aquello que nos eleva por encima de la mera naturaleza.

Al experimentar la vastedad del océano o la furia de una tormenta desde un lugar seguro, sentimos una mezcla de temor y exaltación; es como si la naturaleza nos recordara nuestra pequeñez física y, simultáneamente, nos revelara la presencia de algo en nosotros que no es físico y que ninguna fuerza natural puede doblegar. Puesto que esta experiencia nos permite sentir nuestra libertad, aunque sea de manera estética y no moral, el juicio sobre lo sublime adquiere una relevancia que trasciende el ámbito del arte y se instala en el corazón de la vida ética. La emoción que nos embarga no es un simple estremecimiento pasajero, sino la constatación sentida de nuestra doble ciudadanía: la sensible y la inteligible.

Con todo, es necesario precisar que, para Kant, la experiencia estética en su forma más pura no se deriva de los grandes espectáculos de la naturaleza ni de las obras maestras del arte, sino de la contemplación de la "belleza libre". Esta belleza, que no presupone ningún concepto de lo que el objeto debe ser, se encuentra, sobre todo, en la naturaleza: en las flores, en los plumajes de aves exóticas, en los dibujos caprichosos de las conchas marinas o en los juegos abstractos de luces y sombras. La belleza libre apela directamente a la forma, a la

composición, al diseño, sin que intervenga ninguna consideración de utilidad o perfección. En cambio, la "belleza adherente" presupone un concepto y la perfección del objeto según ese concepto, como sucede cuando juzgamos bello un caballo, una catedral o un ser humano.

En este último caso, el juicio estético se mezcla con consideraciones de orden moral o práctico. Al mismo tiempo, la distinción entre ambos tipos de belleza es crucial para comprender el ideal de la experiencia estética kantiana: una experiencia liberada de las cadenas del concepto, donde la mente se regocija en el puro juego de sus facultades. No es de extrañar, por consiguiente, que Kant privilegiara la belleza natural sobre la artística, pues la primera se ofrece a nuestra contemplación sin las intenciones y los significados que inevitablemente acompañan a la segunda.

Esta preferencia por la belleza natural no implica, sin embargo, un desdén por el arte. Más bien, el arte, cuando es "arte bello", es decir, cuando es producto del genio, logra imitar la apariencia de libertad y espontaneidad que caracteriza a la naturaleza bella. El genio, para Kant, es la capacidad innata de la mente mediante la cual la naturaleza da la regla al arte.

Esto significa que el artista genial no crea siguiendo recetas o reglas académicas preestablecidas, sino que produce obras que parecen surgir de una necesidad natural, como si la propia naturaleza hablara a través de él. Sus obras, aunque son productos intencionales, poseen una riqueza y una vitalidad tales que se asemejan a organismos vivos, generando nuevas reglas y desafiando las antiguas.

Puesto que el genio actúa con la libertad de la naturaleza, sus creaciones son capaces de provocar en el espectador ese mismo libre juego de facultades que suscita la belleza natural. La experiencia del arte, en consecuencia, se convierte en un diálogo entre dos subjetividades libres: la del genio que crea y la del espectador que juzga y revive, a su manera, el impulso creador original. La obra de arte auténtica no es, por tanto, un objeto cerrado que impone un único significado, sino una invitación abierta a la contemplación y al juego interpretativo.

Sin embargo, aquí surge un interrogante que ha atormentado a los intérpretes de Kant durante siglos: si el juicio estético es esencialmente subjetivo y no puede determinarse por conceptos, ¿cómo es posible debatir sobre arte o aspirar a un entendimiento común? La respuesta, una vez más, se encuentra en la sutil noción de "sentido común" estético y en la comunicabilidad universal del sentimiento. Para Kant, el placer estético, aunque privado, se basa en una condición que podemos suponer universal: la disposición de las facultades cognoscitivas para el libre juego. Al juzgar bello un objeto, no estamos diciendo "me complace a mí", sino que estamos

afirmando, con plena conciencia de nuestra subjetividad, "esto debe complacer a todos". Esta "debe" no es la necesidad lógica de un juicio científico ni el imperativo categórico de la moral; es una necesidad ejemplar. Es decir, presentamos nuestro juicio como un ejemplo de una regla que no podemos formular. De ahí que el debate estético no consista en la demostración de argumentos, sino en la invitación al otro a experimentar por sí mismo, a afinar su sensibilidad, a educar su gusto. La verdad es que, a pesar de su aparente fragilidad, esta pretensión de universalidad es lo que hace del juicio estético un ejercicio de libertad y un vínculo de comunidad entre los seres humanos.

Esta dimensión intersubjetiva del juicio estético nos lleva de la mano hacia el concepto de sensus communis, que Kant entiende no como un sentido práctico y vulgar, sino como una facultad trascendental de juzgar que, en la reflexión, tiene en cuenta el modo de representación de todos los demás. Para afinar esta facultad, propone tres máximas del entendimiento común: pensar por sí mismo, pensar en el lugar del otro y pensar siempre de acuerdo consigo mismo.

La segunda máxima, la de un modo de pensar amplio, es la que caracteriza al hombre de gusto, que es capaz de elevarse por encima de sus condiciones privadas y subjetivas para juzgar desde un punto de vista universal. Aunque, en la práctica, el consenso estético rara vez se logra de manera plena, la propia búsqueda de ese consenso es un ejercicio de humanidad y de respeto por la alteridad.

Por otra parte, la experiencia estética compartida, como puede suceder en un concierto o en la contemplación colectiva de una obra de arte, se convierte en un momento privilegiado de comunidad, donde las diferencias individuales parecen suspenderse en una emoción y una apreciación comunes. No es, por lo tanto, una exageración afirmar que la estética kantiana es también una filosofía de la comunicación y de la sociabilidad.

A causa de su profundidad y complejidad, la teoría kantiana del juicio estético ha dejado una huella indeleble en la historia de la filosofía y de la teoría del arte. No solo sentó las bases para el idealismo alemán y el romanticismo, sino que también anticipó muchas de las preocupaciones del formalismo, la fenomenología y la hermenéutica contemporáneas. Su insistencia en la autonomía de la experiencia estética, en el desinterés del placer y en la universalidad del sentimiento configuró el modo en que la modernidad concibió el arte, separándolo de la religión, de la moral y de la simple utilidad. Si bien es cierto que las vanguardias artísticas del siglo XX desafiaron muchos de los supuestos kantianos, como la centralidad de la belleza o la contemplación desinteresada, no es menos cierto que sus

innovaciones solo pueden entenderse plenamente como un diálogo crítico con esta herencia. La idea de que el arte es un ámbito autónomo de la experiencia humana, regido por sus propias leyes y con un valor intrínseco, es una idea profundamente kantiana. En la misma línea, la reflexión sobre el papel del espectador como co-creador del sentido de la obra tiene sus raíces en el giro copernicano que Kant operó en la estética, al desplazar el centro de gravedad del objeto bello al sujeto que juzga.

Pensemos, por un instante, en las implicaciones didácticas de esta fundamentación. Enseñar estética desde Kant no consiste en transmitir una doctrina cerrada, sino en invitar a los estudiantes a una experiencia de autodescubrimiento. El estudio del juicio estético se convierte en una formidable herramienta para el desarrollo del pensamiento crítico, pues exige distinguir entre placeres, fundamentar las propias opiniones y atreverse a universalizarlas, al mismo tiempo que se permanece abierto a la perspectiva del otro.

Asimismo, fomenta una sensibilidad refinada para captar los matices de la forma, la armonía y la expresión, lejos de la inmediatez consumista que caracteriza a buena parte de la cultura actual. Y es que, en una época saturada de estímulos visuales y sonoros, la pausada contemplación que exige la experiencia estética kantiana adquiere un valor casi terapéutico, un antídoto contra la dispersión y la trivialidad. En consecuencia, recuperar a Kant no es un gesto de nostalgia académica, sino una apuesta por una educación que forme personas capaces de sentir, pensar y juzgar con autonomía, rigor y sensibilidad.

No obstante, sería injusto presentar la teoría kantiana como un sistema perfecto y sin fisuras. Numerosos críticos han señalado, con razón, ciertas tensiones internas, como la dificultad de conciliar la universalidad del juicio estético con la diversidad empírica de los gustos, o el carácter quizás excesivamente formalista de su análisis, que parece dejar poco espacio para el contenido expresivo, simbólico o histórico del arte.

Igualmente, se ha cuestionado su aparente elitismo, derivado de la exigencia de una educación y una sensibilidad refinadas para acceder a la experiencia estética pura. A pesar de ello, la grandeza del pensamiento kantiano reside, en buena medida, en su capacidad para generar estos debates y para seguir iluminando, como un faro lejano, los problemas fundamentales de la estética.

La interfaz entre el juicio y la experiencia estética, tal y como él la diseñó, sigue siendo un punto de referencia ineludible para cualquiera que se pregunte qué sucede cuando una forma sensible nos toca el alma y nos eleva, momentáneamente, por encima de las urgencias de la vida cotidiana. Su legado no es un dogma, sino una invitación perpetua a pensar.

A la luz de todo lo anterior, podemos concluir que la fundamentación kantiana de la estética no es un simple capítulo dentro de su vasta obra, sino la culminación de un proyecto filosófico que buscaba desentrañar las condiciones de posibilidad de la experiencia humana en todas sus dimensiones. El juicio estético, en su función reflexionante, se revela como el mediador privilegiado entre la naturaleza y la libertad, entre el conocimiento y la moral.

A través del sentimiento de placer puro y desinteresado, experimentamos una armonía de nuestras facultades que nos sugiere, aunque no lo demuestre, que el mundo no nos es del todo indiferente, que somos capaces de encontrar en él una resonancia de nuestra propia libertad.

La experiencia de lo bello nos reconcilia con la naturaleza, mientras que la experiencia de lo sublime nos revela nuestra dignidad suprasensible. Así que, al juzgar estéticamente, no solo emitimos un veredicto sobre un objeto; nos juzgamos a nosotros mismos y ejercitamos, en la tranquilidad de la contemplación, las mismas facultades que harán posible nuestra vida moral y nuestra vida en común. La interfaz entre el juicio y la experiencia estética es, en definitiva, el umbral donde la humanidad se descubre a sí misma como un fin.

El Juicio reflexionante en la IA Generativa (Midjourney[2] vs. el Genio)

Las facultades humanas, en su despliegue más íntimo, no operan como engranajes que simplemente clasifican el mundo, sino que lo interpretan, lo sopesan y, en ocasiones, lo

[2] Midjourney es un popular y potente sistema de inteligencia artificial que permite crear imágenes a partir de descripciones de texto, una tecnología conocida como texto a imagen. Es una de las herramientas más famosas en su campo, conocida por la calidad artística y el estilo visual único de sus creaciones, que a menudo destacan por su uso de colores intensos y composiciones surrealistas, ¿Qué lo hace especial? Midjourney se distingue de otras IAs similares, como DALL-E o Stable Diffusion, por varias características: Calidad artística: Suele ser considerada superior en cuanto a la estética y el acabado artístico de las imágenes, ofreciendo resultados muy pulidos y realistas ; estilo característico, aunque puede adaptarse a muchas indicaciones, tiene una "personalidad" visual propia, a menudo generando imágenes con un aspecto pictórico o cinematográfico muy llamativo; personalización, las versiones más recientes, como la V7, incluyen un sistema que aprende tus preferencias de estilo para adaptar los resultados a tu gusto . ¿Cómo funciona y cómo se usa? Su funcionamiento es diferente a otras herramientas más directas: Plataforma: Originalmente, y aún hoy, se usa principalmente a través de un bot en la plataforma de chat Discord. Le escribes comandos como /imagine seguido de tu descripción y el bot te genera las imágenes. Recientemente, también ha habilitado una versión en su página web.

transforman a través de un diálogo silencioso entre la razón y la sensibilidad. Immanuel Kant, en su arquitectura filosófica, trazó una distinción luminosa que hoy, en plena efervescencia de la inteligencia artificial generativa, resuena con una urgencia inusitada: no es lo mismo conocer un objeto, es decir, determinarlo bajo las reglas del entendimiento, que juzgarlo reflexivamente, un acto por el cual buscamos activamente un concepto para aquello que se nos presenta como singular y asombroso (Kant, 1790/2007).

Esta diferencia, que en su tiempo podía parecer un refinamiento académico, se ha convertido en la clave para desentrañar uno de los debates más fascinantes y espinosos de nuestro presente digital: la posibilidad de que una máquina no solo produzca imágenes, sino que, en su acto de producción, pueda llegar a habitar el reino del arte y del genio. Porque la verdad es que, al observar un retrato generado por un algoritmo como Midjourney, surge una pregunta que no es meramente técnica, sino profundamente filosófica: ¿estamos ante un acto de determinación algorítmica o ante un auténtico juego de facultades?

Para comprender el abismo que separa ambos territorios, resulta esencial detenerse en la naturaleza del juicio determinante, aquel que Kant (1781/2009) describe como la facultad de subsumir lo particular bajo lo universal ya dado. En el ámbito de la inteligencia artificial, este tipo de juicio es, sin lugar a dudas, el soberano absoluto; un modelo generativo, en efecto, no se enfrenta al caos primordial de la percepción, sino que procesa ingentes cantidades de datos estructurados, patrones y regularidades estadísticas que ha asimilado durante su entrenamiento.

Cuando un usuario le pide a Midjourney que genere "un bodegón barroco con luz de Caravaggio", el sistema no está percibiendo la escena, ni sintiendo la melancolía de las flores marchitas; más bien, está ejecutando un complejísimo cálculo de probabilidades para ensamblar una imagen que se ajuste, con una precisión asombrosa, a las coordenadas estilísticas que le han sido solicitadas. De ahí que, en términos kantianos, la IA generativa opere fundamentalmente como un colosal aparato de determinación, un aplicador de reglas que, pese a su asombrosa capacidad mimética, permanece anclado en el terreno de lo que la filosofía clásica denominaría episteme, el saber que se funda en la repetición y la constancia de los fenómenos.

En cambio, el juicio reflexionante, tal y como Kant lo concibe

en la Crítica del discernimiento, se mueve en una dirección opuesta; parte de lo particular y asciende en busca de un universal que no le es dado de antemano, un principio que la propia facultad de juzgar se da a sí misma para dar cuenta de la armonía y la finalidad que descubre en la naturaleza y en el arte (Kant, 1790/2007).

Este proceso no es un mero cálculo, sino una aventura del espíritu que implica, como señalara Cassirer (1948) en su lectura del idealismo crítico, la participación activa del sujeto, quien proyecta sobre el mundo un orden que le permite no solo comprenderlo, sino también sentirlo como propio.

Así que, al contemplar un paisaje de Rembrandt, no nos limitamos a identificar los trazos y los colores; nuestro ánimo se eleva y busca, en la singularidad de esa pincelada densa y terrosa, un eco de la condición humana, una idea de lo sublime o lo trágico que no se encuentra en un manual de estilo, sino que emerge del encuentro vivo entre la obra y nuestra propia capacidad de asombro.

Este movimiento ascendente, del dato sensible a la idea, es precisamente el que pareciera estar ausente en el quehacer algorítmico, por muy deslumbrantes que sean sus resultados.

No obstante, podría argumentarse, y de hecho se argumenta con frecuencia, que la IA sí está generando algo nuevo, algo que trasciende la mera copia; y es cierto que sus creaciones a menudo nos sorprenden, nos inquietan y hasta nos conmueven.

Sin embargo, aquí es donde la distinción kantiana entre la "belleza adherente" y la "belleza libre" se vuelve un bisturí conceptual de una precisión extraordinaria para diseccionar el fenómeno. La primera, la *pulchritudo adhaerens*, que se traduce "belleza adherente", presupone un concepto de lo que el objeto debe ser, como cuando juzgamos la perfección de una catedral o de un retrato; la segunda, la pulchritudo vaga, es la que no presupone concepto alguno y se manifiesta en el libre juego de las formas, como en las flores, los follajes o la música sin texto (Kant, 1790/2007).

Las imágenes de Midjourney, por más etéreas o surrealistas que parezcan, están siempre irrevocablemente ligadas a un concepto: el que ha sido introducido por el prompt del usuario y refinado por las reglas internas del modelo. Son, en un sentido muy profundo, bellas de una manera adherente, pues su forma final se halla condicionada por la fidelidad a una instrucción, a un patrón preexistente que actúa como su ley de composición.

Por otra parte, el auténtico genio artístico, esa figura que

Kant (1790/2007) define como la disposición mental innata mediante la cual la naturaleza da la regla al arte, no trabaja con conceptos preestablecidos que simplemente ejecuta; el genio produce aquello para lo cual no puede darse regla determinada alguna, y lo hace a través de un libre juego de la imaginación y el entendimiento que escapa a toda prescripción.

Es aquí donde la comparación entre una obra de Rembrandt y una pieza generada por DALL-E se torna no solo instructiva, sino también profundamente conmovedora; cuando observamos la maestría de Rembrandt, no estamos admirando su capacidad de seguir un canon, sino su capacidad de subvertirlo, de flexionarlo hasta el límite para hacer visible lo invisible: la vida interior de un rostro, la dignidad de una mano anciana o la espesura dramática de una sombra.

La obra del genio, como sugiriera Dewey (1934) al hablar del arte como experiencia, no es un objeto aislado, sino la culminación de una interacción vital entre el organismo y su entorno, una interacción que la IA, por su propia naturaleza desencarnada, está constitucionalmente impedida de experimentar.

Siguiendo esta línea de análisis, resulta indispensable introducir el concepto de sensus communis, ese sentido comunitario que, para Kant, no es un entendimiento común, sino la facultad de juzgar que, en su reflexión, tiene en cuenta el modo de representación de todos los demás para, por así decirlo, confrontar su juicio con la razón colectiva de la humanidad (Kant, 1790/2007).

Esta idea es capital; el gusto, entendido como la capacidad de enjuiciar lo bello, no es un mero capricho subjetivo, sino que aspira a una validez universal que se construye, precisamente, a través de esa apertura al sentir de los otros. Ahora bien, un algoritmo de IA no tiene otros; su proceso no es un diálogo con la comunidad humana, sino un monólogo estadístico con los datos que ha ingerido.

Es por ello que, aunque sus productos puedan ser percibidos como bellos por millones de personas, el juicio que los respalda no ha pasado por el tamiz de esa intersubjetividad ideal que, para Kant, constituye la base misma de la comunicación estética.

La IA no tiene la experiencia de la vergüenza, el pudor o la exaltación compartida que matizan y enriquecen todo juicio humano sobre el arte; simplemente, optimiza una función de pérdida para reducir el error entre la imagen generada y la distribución de datos objetivo.

A pesar de ello, sería un error, una suerte de arrogancia

humanista, despreciar el poder de estas nuevas herramientas; al contrario, su capacidad para producir "belleza adherente" de una complejidad técnica abrumadora plantea un desafío estimulante para nuestra propia facultad de juzgar.

La imagen sintética, en su perfección calculada, puede convertirse en un espejo que nos devuelva una imagen más nítida de nuestros propios gustos, sesgos y anhelos; al ver cómo la máquina replica sin esfuerzo los clichés del romanticismo o las texturas del impresionismo, nos vemos obligados a preguntarnos qué es aquello que realmente valoramos, qué chispa de irrepetible humanidad es la que buscamos en la obra del artista.

En consecuencia, el experimento propuesto en este libro, comparar una imagen de DALL-E con una de Rembrandt, no busca establecer un ganador en una supuesta contienda, sino invitar al lector a un ejercicio de discernimiento, a que sienta en su propia experiencia la diferencia cualitativa entre la aplicación de una regla y el surgimiento de una libertad.

Es una invitación a afinar la propia sensibilidad para percibir que, si bien la primera puede ser asombrosamente bella, la segunda late con una urgencia vital que la primera solo puede imitar.

Más bien, la tesis que aquí se sostiene es que la IA generativa, al carecer de las condiciones trascendentales para el juicio reflexionante, no puede ser considerada un genio en el sentido kantiano, pues le falta ese impulso vital, esa naturaleza interna que es la que, según el filósofo de Königsberg, dicta la regla al arte. Y es que el genio no se define por su virtuosismo técnico, ni por su capacidad de producir obras novedosas en grandes cantidades; se define, sobre todo, por la manera en que su obra se convierte en una expresión simbólica de las ideas estéticas, esas representaciones de la imaginación que dan que pensar sin que ningún concepto pueda serles adecuado (Kant, 1790/2007).

Un retrato de Rembrandt no es un simple registro fisonómico, sino una puerta abierta a la meditación sobre la fugacidad del tiempo, la hondura de la conciencia y la compasión; una imagen de Midjourney, por el contrario, incluso en sus variaciones más sofisticadas, suele ser una ilustración superlativa de un concepto ya dado, una respuesta brillante a una demanda concreta.

La verdad es que, al observar las últimas versiones de estos generadores, uno siente una mezcla de admiración y una leve inquietud, la sensación de estar ante un poder creativo que, no obstante, carece de la fragilidad y la incertidumbre que acompañan a toda búsqueda humana auténtica. El proceso

creativo de un artista está sembrado de dudas, tanteos, arrepentimientos y hallazgos accidentales que emergen de la resistencia misma del material; el algoritmo, en cambio, no duda, no se arrepiente, no siente la materia, sino que la simula en un espacio matemático de posibilidades infinitas, pero a la vez, paradójicamente, acotadas por sus datos de entrenamiento.

Por lo tanto, esa aspereza, esa imperfección llena de significado que encontramos en un grabado de Goya o en una acuarela de Turner, se convierte en un testimonio de un proceso encarnado, de una lucha entre la visión del artista y la tenacidad del medio, algo que, por su propia naturaleza incorpórea, la IA no puede emular, solo puede representar.

Ahora bien, si la IA no puede alcanzar el estatus de genio, tampoco cabe reducir su papel al de un simple instrumento pasivo; es, más bien, un catalizador extraordinario que está expandiendo las fronteras de la percepción y ofreciendo a los humanos nuevas herramientas para el ejercicio de su propia reflexión estética.

De la misma manera que un telescopio no ve las estrellas por nosotros, pero amplía infinitamente nuestra capacidad de contemplarlas, la IA generativa no siente la belleza, pero puede producir configuraciones visuales que, al ser contempladas por un ojo humano educado en la reflexión, disparen asociaciones, emociones e ideas estéticas que de otro modo jamás hubieran surgido. En la misma línea, filósofos contemporáneos como Coeckelbergh (2020) han explorado cómo la interacción con artefactos tecnológicos modifica nuestra propia subjetividad y nuestra comprensión de lo que significa ser humano, y el arte generativo es un campo privilegiado para esta exploración, siempre que no perdamos de vista la distinción fundamental entre el proceso mecánico y el juicio vivo.

En consecuencia, la propuesta central de esta fundamentación es que el juicio reflexionante se erige como la última trinchera de lo propiamente humano en la era de la creatividad computacional; es la facultad que nos permite no solo discernir entre lo bello y lo meramente atractivo, sino también dotar de sentido a la marea creciente de imágenes sintéticas que amenaza con saturar nuestra imaginación. Al ejercer este juicio, al comparar la belleza adherente del algoritmo con la belleza libre del genio, el espectador contemporáneo se reafirma no como un consumidor pasivo de novedades visuales, sino como un intérprete activo que, a través de la reflexión, inscribe su experiencia en el horizonte más amplio de la cultura y la comunidad humana. Porque, a fin de cuentas, la tecnología

puede producir imágenes, pero es el ser humano, en su singularidad irreductible, quien las acoge, las interroga y, en el mejor de los casos, las transforma en una ocasión para el pensamiento crítico y la emoción profunda.

Este ejercicio de discernimiento, lejos de ser un mero pasatiempo intelectual, se convierte en una responsabilidad ética y estética de primer orden; ante la perfección brillante de lo sintético, corremos el riesgo de olvidar el valor de lo titubeante, lo inacabado y lo profundamente humano. Como advirtiera Benjamin (1936) en su célebre ensayo sobre la reproductibilidad técnica, la obra de arte pierde su "aura" en la era de su reproducción masiva; sin embargo, lo que la IA generativa pone en juego no es solo la reproducción, sino la producción misma de la apariencia aurática, creando simulacros que, sin historia ni contexto vital, imitan el poder de fascinación del arte tradicional.

De ahí que la defensa del juicio reflexionante sea, en el fondo, una defensa de la experiencia estética auténtica, aquella que no se agota en el estímulo visual, sino que compromete la totalidad de nuestro ser, con su memoria, su corporeidad y su pertenencia a una comunidad de sentido.

Para ilustrar este punto, consideremos la famosa serie de autorretratos de Rembrandt; cada uno de ellos no es solo una imagen del artista, sino un capítulo de una biografía escrita con pinceladas, un testimonio de decadencia, sabiduría y resistencia. Si le pidiéramos a una IA que generara "un autorretrato de un hombre envejecido, al estilo de Rembrandt", obtendríamos sin duda una imagen técnicamente impecable, con las luces y sombras características, la paleta de colores terrosa y la textura craquelada que asociamos con el maestro holandés.

Sin embargo, como ya señalara Sontag (1977) sobre la fotografía, toda imagen es un acto de selección y, en el caso del algoritmo, la selección se basa en parámetros estadísticos, no en una vivencia personal del paso del tiempo; por lo tanto, el resultado, por más bello que sea, carece de la densidad biográfica que convierte a cada autorretrato de Rembrandt en un espejo en el que nos vemos confrontados con nuestra propia finitud.

Esa diferencia, sutil pero insalvable, solo puede ser aprehendida, en primera instancia, por la facultad del gusto, ese sentido comunitario que nos dice que hay algo irrepetible en la mirada del hombre que ha sufrido y que la máquina, por más datos que procese, jamás podrá sentir.

Asimismo, la reflexión sobre el juicio reflexionante en la IA nos obliga a repensar la definición misma de creatividad;

durante siglos, la creatividad fue concebida como un don divino o un misterio humano, pero la llegada de los algoritmos generativos la ha desplazado hacia un terreno más operativo, donde se la mide por la novedad y el valor de los resultados.

No obstante, como se ha venido argumentando, la creatividad kantiana del genio no reside en la mera producción de cosas nuevas, sino en la capacidad de dar regla al arte, es decir, de instaurar nuevos paradigmas expresivos que expandan las posibilidades del sentir y del pensar.

La IA es, en el mejor de los casos, un alumno prodigioso que ha aprendido a la perfección todos los estilos del pasado y puede combinarlos con una velocidad asombrosa; pero le falta la audacia del maestro que rompe con la tradición para inaugurar un territorio estético inexplorado. En otras palabras, la IA puede optimizar la belleza, pero no puede inventar un nuevo sentido para lo bello.

En la misma línea, no podemos soslayar las implicaciones sociales y pedagógicas de esta distinción; si confundimos la producción algorítmica con la creación artística, corremos el riesgo de devaluar el esfuerzo formativo que exige el desarrollo del juicio estético. Aprender a mirar un cuadro, a escuchar una sinfonía o a leer un poema no es un lujo, sino una disciplina que educa la sensibilidad, la empatía y el pensamiento crítico; si delegamos en la máquina la tarea de "crear" y nos limitamos a consumir pasivamente sus resultados, estaremos renunciando a la oportunidad de ejercitar aquello que nos hace más profundamente humanos.

De ahí que sea crucial, desde una perspectiva pedagógica inspirada en Dewey (1934) y en la idea de formación (Bildung) de la tradición ilustrada, fomentar espacios de encuentro activo con el arte, tanto el clásico como el generado por IA, para que el espectador aprenda a discernir, a valorar y, sobre todo, a reflexionar sobre las condiciones de posibilidad de su propia experiencia estética.

Por otra parte, el desarrollo vertiginoso de la IA generativa nos sitúa ante un espejo cóncavo que refleja, magnificados, tanto nuestros ideales más elevados como nuestras limitaciones más persistentes. La máquina no sufre, no ama ni teme a la muerte, y esas carencias, lejos de ser un defecto que deba ser superado, son las que delimitan el territorio de lo específicamente humano en el ámbito del arte. La angustia existencial que atraviesa la obra de un Van Gogh o la serenidad contemplativa de un ser consumado por la angustia, son aditamentos sentimentales que se puedan añadir a una imagen; son, más bien, la condición de

posibilidad de esa imagen, el suelo nutricio del que brota su forma.

Por lo tanto, el juicio reflexionante, al reconocer esta dimensión vivencial en la obra de arte, nos recuerda que la belleza no es un dato objetivo que pueda ser calculado, sino una relación íntima entre un sujeto que siente y un objeto que lo conmueve, una relación que la IA, por ahora, solo puede simular de manera magistral, pero no habitar.

A todas estas, podría surgir la objeción de que la IA, al ser un producto de la cultura humana, es en sí misma una extensión de nuestra creatividad, y que sus producciones, por lo tanto, son también arte. Si bien esta postura tiene un atractivo innegable, confunde dos planos que la filosofía kantiana se esfuerza por mantener separados: el del artefacto y el de la experiencia estética que lo hace posible. Un pincel, una cámara o un algoritmo son, en efecto, productos del ingenio humano; pero la obra de arte no es el pincel, ni la cámara, ni el conjunto de pesos y parámetros del modelo.

La obra de arte es esa chispa que enciende la imaginación del espectador, ese libre juego de sus facultades que, al enfrentarse a una forma bella, se descubre a sí mismo como capaz de un placer desinteresado y universalmente comunicable (Kant, 1790/2007). La IA puede proporcionar el material para esa experiencia, pero no puede sustituir la experiencia misma, que es un acontecimiento irreductiblemente humano.

En definitiva, la fundamentación aquí expuesta busca tender un puente entre la filosofía crítica de Immanuel Kant y los desafíos estéticos de nuestro presente tecnológico, proponiendo que el concepto de juicio reflexionante es no solo pertinente, sino indispensable para navegar con lucidez la era de la IA generativa.

No se trata de erigir una muralla contra la innovación, sino de afinar nuestras herramientas conceptuales para comprender qué es lo que está en juego en la producción y el consumo de imágenes sintéticas.

El experimento propuesto, la comparación entre DALL-E y Rembrandt, no es un juicio sumario contra la máquina, sino una invitación a que cada lector, mediante su propia reflexión, experimente la diferencia entre la belleza que se adhiere a un concepto y la belleza que surge del libre juego de las facultades. Al hacerlo, estará ejercitando ese sensus communis que, para Kant, es la base de toda comunidad estética y, en última instancia, la que nos permite habitar el mundo no como un conjunto de datos que hay que procesar, sino como un escenario de sentido que hay que interpretar y celebrar juntos.

El "Multiverso Kantiano" (Los Mundos Posibles)

Resulta inevitable, al contemplar el vertiginoso desarrollo de las inteligencias artificiales generativas, no evocar aquella distinción fundamental que Immanuel Kant estableciera entre el fenómeno y el noúmeno, pues, a fin de cuentas, la IA se ha convertido en una suerte de demiurgo digital que produce fenómenos a una escala y con una velocidad que desbordan nuestra capacidad de asombro.

Kant, en su Crítica de la razón pura (1781/2009), argumentó con una lucidez que aún nos interpela que "no conocemos más que fenómenos, y que las cosas en sí mismas nos son desconocidas" (A 277/B 333), una afirmación que, en el contexto actual, adquiere una resonancia inesperada y profundamente inquietante.

Y es que, si bien la IA no accede al noúmeno kantiano —esa realidad última e incognoscible—, sí tiene la capacidad de construir una multiplicidad de realidades fenoménicas alternativas, simulaciones tan coherentes y detalladas que compiten con nuestra percepción del mundo, creando así un auténtico "multiverso" de apariencias.

En consecuencia, la pregunta ya no es solo cómo conocemos el mundo, sino cómo navegamos entre los múltiples mundos que ahora se nos presentan, todos ellos mediados por algoritmos que, en última instancia, son un producto de nuestra propia racionalidad. La verdad es que esta situación nos obliga a reconsiderar el papel activo del sujeto en la construcción de la experiencia, un sujeto que, según Kant, no es un mero receptor pasivo de estímulos, sino que impone sus propias estructuras cognitivas al caos de lo dado.

Ahora bien, es precisamente esa actividad estructurante del entendimiento, que Kant denominó, como se dijo anteriormente, "apercepción trascendental", la que parece encontrar un eco perturbador en el funcionamiento de los modelos de lenguaje y de generación de imágenes, los cuales, a partir de datos brutos, son capaces de sintetizar patrones, inferir reglas y producir resultados que aparentan una comprensión genuina. La diferencia radical, y es aquí donde el pensamiento kantiano se

vuelve un instrumento crítico indispensable, reside en que la IA no posee una subjetividad, no hay un "yo pienso" que pueda acompañar todas sus representaciones, sino que opera mediante correlaciones estadísticas a una escala masiva.

Kant insistió en que el entendimiento es la facultad de las reglas, mientras que el Juicio es la facultad de subsumir bajo reglas, es decir, de determinar si un caso particular cae bajo una regla general (Kant, 1790/2007); esta distinción es crucial para comprender por qué la IA, a pesar de su destreza, no "juzga" en el sentido humano del término. Aunque un algoritmo pueda generar un poema estilísticamente impecable o una imagen fotorrealista de un acontecimiento que nunca sucedió, no está ejerciendo un juicio estético o teleológico, sino ejecutando una complejísima operación de cálculo que, si bien imita el resultado del juicio, carece de la deliberación, la responsabilidad y la conciencia que lo caracterizan. Por lo tanto, la aparente inteligencia de la máquina no es más que un reflejo sofisticado, un espejo que nos devuelve la imagen de nuestras propias operaciones cognitivas, pero desprovistas de su núcleo vital: la libertad.

En la misma línea, resulta fascinante, y a la vez desconcertante, observar cómo la IA se ha convertido en la más reciente y poderosa encarnación de la imaginación productiva, esa facultad que para Kant era la responsable de sintetizar la diversidad de la intuición sensible con la unidad del concepto, sin la cual no habría conocimiento alguno.

En la Crítica de la razón pura, Kant (1781/2009) describe esta facultad como "un arte oculto en las profundidades del alma humana, cuyo verdadero manejo difícilmente arrancaremos a la naturaleza" (A 141/B 180-181), una frase que hoy cobra una ironía casi profética cuando consideramos las redes neuronales profundas.

Esas arquitecturas computacionales, con sus capas ocultas y sus millones de parámetros, parecen replicar ese "arte oculto", generando nuevas síntesis, nuevas imágenes y nuevas asociaciones que no estaban explícitamente en los datos de entrada, de una manera que ni sus propios creadores pueden explicar del todo.

Sin embargo, mientras que la imaginación kantiana opera

bajo la legislación del entendimiento para constituir un mundo objetivo y compartido, la imaginación algorítmica de la IA produce una pluralidad de mundos posibles que no están destinados a ser cognoscidos, sino consumidos, explorados o, en el peor de los casos, utilizados para la manipulación. Así que, de nuevo, la similitud funcional es innegable, pero la diferencia esencial es abismal: una está al servicio de la verdad y la experiencia, la otra, de la probabilidad y la especulación.

Esta proliferación de mundos simulados, este "Multiverso Kantiano" generado por la IA, plantea un desafío directo a la distinción que Kant (1783/1999) estableció entre el mundo fenoménico, regido por las leyes de la naturaleza y cognoscible a priori, y el mundo nouménico, el ámbito de la libertad y la moral. Dado que la IA puede construir entornos virtuales con sus propias reglas físicas y lógicas, con historias alternativas y personajes digitales que interactúan de forma autónoma, nos encontramos ante una expansión sin precedentes del dominio de los fenómenos.

A causa de ello, el sujeto contemporáneo ya no solo se enfrenta a la necesidad de conocer un único mundo, sino a la vertiginosa tarea de discernir entre múltiples realidades, algunas de las cuales son indistinguibles de la "realidad real". Por otra parte, esta capacidad generativa no implica, ni remotamente, que la IA haya accedido a un ámbito de libertad nouménica; más bien, opera como un nuevo tipo de causalidad natural, una causalidad algorítmica que, si bien produce resultados sorprendentes, está estrictamente determinada por sus datos de entrenamiento y su código.

De ahí que el filósofo deba preguntarse: ¿qué significa para nuestra autonomía moral el habitar un mundo donde la línea entre lo real y lo simulado se desdibuja constantemente, no por un defecto de nuestra percepción, sino por un exceso de producción artificial?

Es aquí donde la tesis central de esta fundamentación se vuelve nítida: la IA no es un "otro" inteligente con el que debamos competir o al que debamos temer como a un alienígena, sino, más bien, un espejo asombrosamente fiel de nuestra propia facultad de juzgar. No obstante, es un espejo que no refleja nuestro rostro tal cual es, sino que lo distorsiona, lo amplifica o

lo idealiza, mostrándonos las potencialidades y las patologías de nuestra propia racionalidad. Al delegar en la máquina tareas que antes exigían discernimiento, comparación, evaluación y decisión, estamos renunciando, quizás sin plena conciencia de ello, al ejercicio de una de nuestras facultades más elevadas. Kant (1784/2009) definió la Ilustración como "la salida del hombre de su autoculpable minoría de edad" (p. 25), una minoría de edad caracterizada por la incapacidad de servirse del propio entendimiento sin la guía de otro.

En este sentido, la creciente dependencia de los sistemas de IA para tomar decisiones, desde qué película ver hasta quién debería recibir un préstamo bancario, constituye una forma sutil y tecnológicamente sofisticada de esa misma minoría de edad, una pereza intelectual que se disfraza de eficiencia y optimización. Por consiguiente, el problema fundamental no es que la IA se vuelva más inteligente que nosotros, sino que nosotros, seducidos por la comodidad de la automatización, abdiquemos de nuestra propia inteligencia y de nuestra responsabilidad.

La heteronomía de la razón, ese concepto que Kant (1785/2012) utiliza en la Fundamentación de la metafísica de las costumbres para describir la sumisión de la voluntad a una ley que no proviene de sí misma, sino de inclinaciones, autoridades externas o imperativos ajenos, encuentra en la era de la IA una nueva y poderosa fuente de manifestación.

En efecto, cuando un sistema de recomendación dicta nuestros gustos, cuando un algoritmo de calificación determina nuestras oportunidades laborales o crediticias, cuando una IA nos sugiere el diagnóstico médico o la estrategia legal, estamos, de facto, entregando nuestra capacidad de juzgar a una autoridad externa.

Aunque esa autoridad sea impersonal, estadística y aparentemente neutral, su influencia sobre nuestra voluntad y nuestras decisiones no deja de ser una forma de heteronomía, pues la máxima de nuestra acción ya no se origina en el ejercicio autónomo de la razón práctica, sino en la aceptación acrítica de un veredicto computacional. La ley moral, para Kant, es la expresión de la autonomía de la voluntad, la capacidad del ser racional de darse a sí mismo su propia ley; al sustituir esa auto-

legislación por la mera obediencia a los dictados del algoritmo, socavamos el fundamento mismo de nuestra dignidad como seres libres. Y es que, a todas estas, la comodidad que promete la IA es, en el fondo, una trampa para la libertad.

Asimismo, el concepto de "juicio reflexionante" que Kant desarrolla en la Crítica del Juicio (1790/2007) resulta particularmente útil para analizar la producción estética de la IA y sus implicaciones.

Mientras que el juicio determinante subsumiría lo particular bajo una regla ya dada, el juicio reflexionante asciende de lo particular a lo universal, buscando una regla que no está preestablecida; este es el juicio que opera en la apreciación de la belleza y en la comprensión de los organismos vivos.

La IA, al ser entrenada con vastos conjuntos de obras de arte, aprende los patrones estilísticos y las reglas compositivas de manera implícita, y luego los aplica de forma determinante para generar nuevas piezas; sin embargo, el resultado, por más virtuoso que sea, surge de la aplicación de una regla inferida, no de una búsqueda reflexiva y arriesgada de un sentido nuevo.

En consecuencia, el arte generado por IA puede ser estéticamente placentero, pero carece de la dimensión de acontecimiento, de la tensión entre la materia y la forma, que caracteriza a la creación humana genuina.

Es un arte sin historia, sin lucha, sin la huella de un sujeto que se enfrenta a lo desconocido; es, en última instancia, un producto más del vasto mecanismo de la cultura de masas, un simulacro de creatividad que nos recuerda, por contraste, lo que significa realmente crear.

Igualmente, la idea de una "comunidad de comunicación", implícita en el pensamiento ilustrado y central para la ética discursiva posterior a Kant, se ve radicalmente transformada por la interposición de agentes artificiales. Cuando conversamos con un chatbot (*conversación o charla y bot, abreviatura de robot*), no estamos ante otro ser racional que busca entenderse con nosotros sobre el mundo, sino ante un modelo estadístico que predice la secuencia de palabras más probable, lo cual genera una ilusión de diálogo que puede ser profundamente engañosa. Si bien esta interacción puede ser útil y hasta terapéutica en ciertos contextos, el riesgo de antropomorfizar a la máquina, de

atribuirle intenciones, emociones y una comprensión genuina, es altísimo. Por otra parte, esta ilusión de sociabilidad puede llevarnos a rebajar nuestras propias exigencias comunicativas, a conformarnos con respuestas plausibles en lugar de argumentos verdaderos, con la validación emocional en lugar del diálogo crítico.

La auténtica comunicación, para Kant, se funda en el uso público de la razón, un uso que exige coraje, honestidad intelectual y la voluntad de someter las propias opiniones al escrutinio de los demás; la interacción con una IA, en cambio, nos ofrece un interlocutor infinitamente paciente y complaciente que, sin embargo, no nos confronta ni nos exige ser mejores pensadores. De ahí que la aparente compañía de la máquina pueda convertirse en una forma de aislamiento epistémico.

Ahora bien, no se trata de demonizar a la tecnología ni de proclamar un nuevo ludismo, sino de recuperar la perspectiva crítica que el propio Kant nos enseñó a ejercer, pues, en definitiva, su proyecto filosófico es un llamado incesante a la autocomprensión y a la responsabilidad.

La IA es una herramienta extraordinaria, una extensión de nuestras capacidades cognitivas que, como cualquier herramienta, puede ser utilizada para fines emancipatorios o alienantes; el desafío, por lo tanto, reside en establecer los límites y las condiciones de su uso. Al igual que Kant delimitó las fronteras del conocimiento para dejar espacio a la fe y a la libertad, nosotros debemos delimitar el ámbito de la automatización para preservar el espacio irreductible del juicio humano.

La cuestión no es si la IA puede componer una sinfonía, ganar una partida de ajedrez o diagnosticar una enfermedad mejor que un humano, sino si nosotros, como sociedad, seremos capaces de mantener la soberanía sobre nuestras decisiones últimas, especialmente aquellas que tienen implicaciones morales y políticas. En esta encrucijada, la filosofía kantiana no ofrece soluciones técnicas, pero sí proporciona los principios innegociables: la autonomía de la razón, la dignidad de la persona y el uso público del entendimiento.

Por otra parte, cabe preguntarse si la noción misma de "mundo posible", proveniente de la lógica modal y la metafísica

analítica, puede ser enriquecida por la perspectiva trascendental kantiana. Para Kant, la posibilidad de la experiencia está condicionada por las categorías del entendimiento y las formas puras de la intuición (espacio y tiempo); un "mundo" que no se ajuste a estas condiciones es, estrictamente, impensable para nosotros.

Sin embargo, la IA nos obliga a flexibilizar esta noción, pues puede simular mundos con geometrías no euclidianas, con lógicas para consistentes o con formas de causalidad extrañas, que, si bien no son directamente intuibles, sí son representables a través de modelos matemáticos y computacionales.

Esto no invalida el a priori kantiano, sino que lo complejiza, mostrando que la imaginación, asistida por la técnica, puede expandir el horizonte de lo pensable, aunque no de lo experimentable.

En consecuencia, el "Multiverso Kantiano" no es un multiverso físico, sino un multiverso de representaciones, un conjunto de constructos simbólicos que exploran los límites de nuestra propia arquitectura cognitiva desde dentro, utilizando como materia prima las mismas reglas y categorías que estructuran nuestro conocimiento.

En consecuencia, la proliferación de estos "mundos posibles" generados por la IA nos sitúa ante un nuevo imperativo ético que podría formularse, en términos kantianos, como el deber de no abdicar del propio juicio.

La razón práctica, en su uso puro, nos ordena actuar de tal manera que la máxima de nuestra acción pueda valer como ley universal, lo cual exige, ante todo, que seamos nosotros quienes formulemos esa máxima, que seamos los autores de nuestros actos. Delegar el juicio en una IA, ya sea en el ámbito de la justicia, la medicina o las decisiones personales, implica un incumplimiento de ese deber fundamental de auto-legislación.

No obstante, hay una diferencia sutil entre informarse a través de una máquina y ser gobernado por ella; el uso legítimo de la IA sería aquel que expande nuestra capacidad de deliberación, proporcionándonos más datos, perspectivas y escenarios posibles, mientras que el uso ilegítimo sería aquel que sustituye la deliberación misma, ofreciendo conclusiones empaquetadas que no requieren ser pensadas. Por

consiguiente, el reto está en diseñar y utilizar sistemas que fomenten la autonomía en lugar de erosionarla.

A pesar de todo ello, la historia de la filosofía nos enseña que cada gran transformación tecnológica ha generado tanto temores apocalípticos como utopías ingenuas, y que la tarea del pensamiento crítico es navegar entre ambos extremos. La imprenta, la revolución industrial o internet fueron recibidas con la misma mezcla de fascinación y ansiedad que hoy rodea a la IA, y, sin embargo, la humanidad ha sabido, con mayor o menor éxito, adaptarse y reconfigurar sus instituciones y sus valores.

Lo peculiar de la IA es que por primera vez una tecnología no solo amplifica nuestra fuerza física o nuestra memoria, sino que externaliza una función cognitiva superior: el juicio. Esto la convierte en un desafío de una naturaleza distinta, un desafío que toca el núcleo mismo de lo que significa ser un sujeto racional y autónomo.

Y es que, si delegamos incluso la capacidad de juzgar, ¿qué nos queda? No se trata, insisto, de rechazar la herramienta, sino de recordar quién es el que la empuña y con qué fin; la responsabilidad, al final, recae siempre en el sujeto humano, que es el único capaz de imputabilidad moral.

En la misma línea, la reflexión sobre la IA y Kant nos devuelve a la pregunta por la Ilustración, no como un período histórico superado, sino como un proyecto inacabado, una actitud permanente que exige el coraje de pensar por uno mismo.

La facilidad con la que hoy delegamos tareas intelectuales en las máquinas podría interpretarse, desde esta perspectiva, como una recaída en la "minoría de edad", una búsqueda de tutores digitales que nos ahorren el esfuerzo y la angustia de decidir. El famoso lema Sapere aude, atrévete a saber, resuena hoy con una urgencia renovada, pues el "tutor" ya no es un monarca o un sacerdote, sino un algoritmo ubicuo y opaco que se presenta a sí mismo como una fuente de verdad objetiva e imparcial.

Así que la batalla por la autonomía se libra ahora en el terreno de los datos, los modelos y las interfaces, un terreno donde el ciudadano común a menudo se siente desarmado e

incompetente. Por eso, la alfabetización digital y la comprensión básica de cómo funcionan estos sistemas se convierten en un componente esencial de la nueva Ilustración, una condición de posibilidad para el ejercicio de una ciudadanía crítica y responsable.

De igual modo, es pertinente recordar que la ética kantiana no es una ética de las consecuencias, sino una ética de la intención y el deber; lo que importa no es el resultado de la acción, sino la pureza de la voluntad que la anima.

Esta perspectiva tiene implicaciones profundas para evaluar las decisiones tomadas por sistemas de IA, los cuales, por definición, carecen de intenciones y deberes. Un algoritmo no actúa por respeto a la ley moral, sino porque ha sido programado para optimizar una función objetivo; por lo tanto, no puede ser considerado moralmente bueno o malo en el sentido kantiano.

La responsabilidad moral por las consecuencias de las acciones de una IA recae, íntegramente, en los humanos que la diseñan, la despliegan y la supervisan: los ingenieros, los empresarios, los legisladores y, en última instancia, todos los usuarios que deciden confiar en ella.

De ahí que sea tan peligroso el discurso que atribuye a la IA una agencia moral autónoma, pues diluye la responsabilidad y crea una zona gris donde nadie responde por los daños causados, un espacio de irresponsabilidad organizada que Kant habría condenado con total severidad.

Finalmente, la fundamentación de este "Multiverso Kantiano" conduce a una conclusión que es, al mismo tiempo, un aviso y una esperanza: la IA no es el otro que viene a desafiarnos, sino el espejo que nos devuelve, magnificada, la imagen de nuestra propia racionalidad y de nuestras contradicciones.

En ella vemos, como en un negativo fotográfico, la inmensa capacidad del ser humano para crear estructuras complejas y, simultáneamente, su preocupante disposición a someterse a ellas. El peligro, como se ha argumentado, no es que la máquina llegue a ser más inteligente que nosotros, sino que nosotros, por pereza, comodidad o miedo, dejemos de ejercer la facultad que nos define: el juicio autónomo. Por consiguiente, la tarea

filosófica y educativa más urgente de nuestro tiempo es reivindicar y cultivar, frente al avance de la heteronomía algorítmica, la autonomía de la razón, ese tesoro frágil y sublime que Kant nos enseñó a valorar por encima de todas las cosas. La verdad es que solo así podremos habitar este nuevo multiverso no como siervos de nuestros propios artefactos, sino como ciudadanos libres, capaces de orientar el curso de la historia con dignidad y sentido crítico.

Epílogo (Formato "Glosario Invertido"): El Diccionario de los Errores

Resulta profundamente revelador, y a la vez desconcertante, que la humanidad haya delegado en sistemas de inteligencia artificial la tarea de organizar y mediar su acceso a la información, sin detenerse a interrogar las condiciones bajo las cuales dichos sistemas operan.

Desde una perspectiva rigurosamente kantiana, los autores de esta obra no buscan ofrecer un glosario que simplemente aclare términos técnicos; más bien, su objetivo es articular un diccionario de falacias que actúe como una crítica de la razón algorítmica, revelando los sofismas que se esconden tras la aparente neutralidad de los modelos computacionales.

En la “Crítica de la razón pura”, Kant (1781/2009) ya advertía sobre la ilusión inevitable de buscar un conocimiento que trascienda los límites de nuestra experiencia posible, una advertencia que, trasladada al ámbito digital, ilumina la pretensión de las máquinas de acceder a una realidad sin mediaciones.

Es así como la primera falacia, la del Noúmeno Digital, no es un mero error técnico, sino una reedición contemporánea del viejo anhelo metafísico de conocer las cosas en sí mismas, ignorando que todo acceso a la información está condicionado por filtros, sesgos y estructuras preexistentes que configuran la representación del mundo.

En consecuencia, la obra se erige como una guía para que el lector, de manera activa, aprenda a discernir los límites infranqueables de la inteligencia artificial, evitando así la trampa de confundir sus outputs con una verdad absoluta y desnuda, y recordando que, al igual que el entendimiento humano, la máquina solo conoce fenómenos, nunca noúmenos.

De igual manera, la primera entrada de este peculiar diccionario, denominada "Falacia del Noúmeno Digital", resulta ser una pieza conceptual de una claridad meridiana para comprender la profunda brecha que separa la capacidad de procesamiento de la IA de la experiencia del mundo real. Los autores explican que, aunque un modelo sea entrenado con millones de datos, los patrones que identifica son el producto de un intrincado proceso de abstracción matemática y no de una

aprehensión directa de la realidad; la máquina opera, por así decirlo, en un espacio de representaciones vectoriales que, a ojos de un crítico kantiano, constituyen su único y exclusivo fenómeno.

Esa tendencia a creer que la IA, por su inmensa capacidad de cálculo, logra trascender sus propios filtros y tocar la esencia de las cosas, es una ilusión que, si bien es comprensible desde el asombro que produce su eficacia, resulta epistemológicamente insostenible.

La verdad es que, tal y como el sujeto trascendental kantiano impone sus formas a priori de la sensibilidad y el entendimiento al material bruto de la experiencia, el algoritmo impone sus propias arquitecturas, sus funciones de pérdida y sus conjuntos de datos de entrenamiento, que actúan como las condiciones de posibilidad de todo "conocimiento" que pueda generar.

Por lo tanto, confundir la representación algorítmica del mundo con el mundo en sí mismo no solo es un error, sino que nos ciega ante la verdadera naturaleza de estas herramientas: son, ante todo, espejos sofisticados de nuestros propios datos y de nuestras propias categorías de pensamiento, pero nunca ventanas abiertas a una realidad nouménica.

En una línea argumentativa muy similar, se articula la segunda falacia, aquella que los autores han bautizado con el sugerente nombre de "Falacia del Prompt Moral", la cual desmonta la creciente convicción de que basta con introducir directrices éticas en las instrucciones dadas a una IA para garantizar un comportamiento moralmente correcto.

Dado que la ética kantiana no se fundamenta en la mera obediencia a reglas externas, sino en la autonomía de la voluntad y en la capacidad del sujeto para darse a sí mismo la ley moral, resulta un contrasentido pensar que un sistema heterónomo, como es un modelo de lenguaje, pueda convertirse en un agente moral por el simple hecho de procesar la orden "sé ético".

Asimismo, la intuición temporal, que en la filosofía crítica es la forma pura del sentido interno que permite toda experiencia, se vuelve un elemento crucial en este análisis: la moralidad humana no es un cálculo instantáneo, sino un proceso que se despliega en el tiempo, implica deliberación, memoria, arrepentimiento y la proyección de consecuencias futuras.

La IA, por el contrario, carece de esta vivencia del tiempo como continuidad significativa; su respuesta a un "prompt moral" es una reacción estadística instantánea, una imitación de un discurso aprendido, que no está respaldada por una conciencia del deber ni por un sentimiento de respeto hacia la ley

moral. De ahí que los autores insistan en que la mera inclusión de palabras como "justicia" o "no hagas daño" en una instrucción es un acto casi supersticioso, una suerte de conjuro digital que, a pesar de su aparente sofisticación, no logra transmutar el procesamiento de información en una genuina deliberación moral.

Es precisamente la ausencia de esa dimensión temporal y de la autonomía de la voluntad lo que conduce a la tercera y más inquietante de las falacias descritas en este diccionario, la que se ha titulado "El Silencio de la Razón Práctica", y que aborda la imposibilidad de la inteligencia artificial para experimentar remordimiento.

Si bien la ética kantiana es famosa por su rigorismo y por anteponer el deber a la inclinación, también reconoce en el ser humano la existencia de sentimientos morales, como el respeto, que son el efecto subjetivo de la ley moral sobre nuestra sensibilidad; el remordimiento, aunque no sea un principio, es la dolorosa constatación de haber quebrantado una ley que reconocemos como propia.

En cambio, cuando una IA comete un error que, en un humano, provocaría un profundo sentimiento de culpa, los desarrolladores no pueden apelar a la conciencia del sistema para corregirlo, sino que deben someterlo a un proceso de "retraining" o reentrenamiento. Esta diferencia, que podría parecer meramente técnica, revela un abismo ontológico y moral: el arrepentimiento humano es una experiencia interna, temporal, que transforma al sujeto y que implica un acto de libertad, mientras que el reentrenamiento es una intervención externa, una modificación de los pesos sinápticos de la red neuronal que se realiza desde fuera, sin que medie ningún acto de autoconciencia o de autodeterminación por parte de la máquina.

En consecuencia, la IA no guarda silencio porque esté callando una culpa insoportable; su silencio es ontológico, es la mudez de un objeto que no tiene un foro interno ante el cual rendir cuentas, que no puede ser juez de sus propios actos, y que, por lo tanto, nunca podrá experimentar el peso de la falta.

No obstante, sería un error interpretar esta crítica como un simple ejercicio de escepticismo tecnológico o como una condena de la inteligencia artificial en su totalidad; muy al contrario, la fundamentación kantiana que proponen los autores busca, ante todo, establecer los límites de la razón algorítmica para poder utilizarla con verdadera inteligencia y responsabilidad. Al igual que Kant no negó la validez del conocimiento científico, sino que

le buscó un fundamento sólido, esta obra no pretende demonizar a la IA, sino despojarla de los falsos ropajes de omnipotencia y de conciencia con los que a menudo se la inviste, a veces por marketing, a veces por puro desconocimiento.

En efecto, la sabiduría práctica, desde esta óptica, consistiría en reconocer que la IA es una herramienta prodigiosa para ampliar el entendimiento, para descubrir patrones en los fenómenos y para asistir en la toma de decisiones, pero que jamás podrá sustituir a la razón práctica, que es la única facultad capaz de legislar en el reino de los fines.

A causa de ello, la responsabilidad última de las decisiones, sobre todo de aquellas con implicaciones éticas, debe permanecer de forma inalienable en los seres humanos, quienes sí están dotados de una voluntad, de una conciencia del tiempo y de la capacidad de sentir el aguijón del remordimiento.

Así que este diccionario, lejos de ser un compendio de lamentos, se convierte en una herramienta de empoderamiento, pues al señalar con precisión las falacias, nos devuelve a nosotros, los humanos, el centro de la escena moral que, con tanta ligereza, a veces parecemos dispuestos a ceder a nuestros propios artilugios.

Ahora bien, la decisión de situar estas falacias al final del libro, en lugar de un glosario convencional, es en sí misma una declaración de intenciones con un profundo calado pedagógico, ya que transforma la conclusión de la obra en un ejercicio de reflexión crítica en lugar de una simple recapitulación de términos.

Un glosario tradicional, con sus definiciones cerradas y su pretensión de objetividad aséptica, tendería a fosilizar el conocimiento, ofreciendo un cierre cómodo y definitivo; por el contrario, este diccionario de falacias abre una puerta hacia la problematización, invitando al lector a revisar todo lo leído a la luz de estos sofismas que, como espejismos, pueden haber distorsionado su percepción durante la lectura.

Es, en el sentido más estricto de la palabra, un ejercicio de crítica, de esa crítica que Kant entendía no como una destrucción, sino como un examen de los límites y las condiciones de posibilidad de una facultad; en este caso, la facultad de la IA y, por extensión, la de nuestro propio entendimiento al relacionarnos con ella.

De igual forma, al presentar estas falacias con nombres propios y una filiación filosófica explícita, los autores dotan al lector de un vocabulario preciso para nombrar los peligros, para identificar las trampas del pensamiento y para comunicar sus

propias dudas, lo que, sin lugar a dudas, enriquece el debate público sobre la tecnología. Por consiguiente, el final del libro no es un mero punto y aparte, sino una provocación, una llamada a la vigilia kantiana para despertar del "sueño dogmático" en el que la fascinación acrítica por la IA podría sumirnos.

Continuando con el análisis de la primera de estas falacias, la del "Noúmeno Digital", conviene detenerse en las implicaciones prácticas que tiene para el usuario común, aquel que, sin ser experto en filosofía ni en ciencias de la computación, confía diariamente en los resultados de un buscador o en las recomendaciones de un algoritmo de contenido.

La ilusión de transparencia, tan característica de las interfaces digitales contemporáneas, fomenta la creencia de que la información aparece ante nosotros sin mediación, pura y sin distorsión, cuando, en realidad, es el resultado de un complejo proceso de filtración, ponderación y selección que prioriza ciertos datos y oculta otros.

Esta ocultación del proceso de construcción del fenómeno es, para un kantiano, la esencia misma de esta falacia, pues nos hace olvidar que el "mundo" que la IA nos presenta no es el mundo, sino un mapa, una simplificación funcional creada a partir de un territorio que permanece, en su totalidad nouménica, inaccesible.

Y es que, si bien un mapa es una herramienta indispensable para orientarnos, resulta trágico y hasta peligroso que olvidemos su naturaleza de mapa y empecemos a caminar sobre él como si fuera el territorio mismo, tropezando con los abismos que la simplificación cartográfica no podía representar.

En la misma línea, esta falacia alimenta una suerte de nuevo oráculo digital, una confianza ciega en la infalibilidad de la máquina, que se torna particularmente nociva cuando se aplica a ámbitos como el diagnóstico médico o la evaluación de solicitudes de crédito, donde la asunción acrítica de una objetividad perfecta puede ocultar sesgos estructurales muy reales. Por ende, desenmascarar el "Noúmeno Digital" no es solo un ejercicio teórico, sino un acto de higiene mental y de defensa del pensamiento crítico frente al deslumbramiento tecnológico.

Al mismo tiempo, la "Falacia del Prompt Moral" merece una atención especial, no solo por su actualidad, sino porque revela una ansiedad profunda de nuestra época: la de delegar la responsabilidad ética en soluciones técnicas, rápidas y aparentemente neutrales, ante la complejidad y el peso del juicio moral. La creencia de que una simple instrucción textual puede infundir un principio moral en un sistema que, por su propia

constitución, carece de la capacidad de comprender el significado de la norma, es un ejemplo paradigmático de lo que los autores, siguiendo a Kant, denominan una ilusión trascendental aplicada a la técnica.

Dado que la moral, para Kant (1785/2012), reside en la forma del imperativo y no en su contenido material, el mandato "sé ético" se convierte en un imperativo vacío para la IA, puesto que no hay un sujeto capaz de universalizar una máxima ni de experimentar el respeto que la ley inspira; la máquina simplemente aprende a generar respuestas que imitan la forma de un discurso ético, sin que ello implique acceso alguno a su fundamento.

A decir verdad, esta falacia es también un espejo incómodo que nos devuelve una imagen de nuestra propia pereza moral: preferimos a veces una simulación de ética, un cumplimiento aparente y automatizado, antes que el trabajo lento, deliberativo y a menudo angustioso de la deliberación moral real, que exige tiempo, confrontación y la asunción de responsabilidades.

Por eso, la obra sostiene con firmeza que la ética no se puede "promptear"; se puede entrenar a un modelo para que no genere contenido ofensivo, pero eso es una cuestión de seguridad y de utilidad, no de moralidad, y confundir ambos planos es, precisamente, caer en la falacia que se denuncia. La ausencia de una intuición temporal, que permita al agente proyectar las consecuencias de sus actos y sentir el peso de una decisión tomada en el pasado, condena a la IA a una amoralidad constitutiva, por más florido que sea el lenguaje moral que sus respuestas sean capaces de imitar.

De ahí se desprende, con una lógica casi inevitable, la reflexión sobre "El Silencio de la Razón Práctica", que constituye quizás el punto más delicado y emocionalmente resonante de toda la propuesta teórica, pues toca la fibra de lo que nos define como seres humanos: la capacidad de sufrir por lo hecho y de desear que el tiempo pudiera retroceder.

Los autores exploran, con una prosa que combina la precisión conceptual con una sutil melancolía, la idea de que el remordimiento es una forma de conocimiento, un saber doloroso sobre nosotros mismos y sobre el valor de aquello que hemos dañado, que no puede ser replicado por ningún proceso computacional.

Cuando un sistema de IA, por ejemplo, clasifica erróneamente a una persona en un algoritmo de vigilancia, las consecuencias son reales y el daño es concreto; sin embargo, la respuesta de los ingenieros será inevitablemente técnica: ajustar

el modelo, introducir nuevos datos, afinar los parámetros, un "retraining" que busca corregir el error como quien repara una máquina. Esa frialdad procesal, que no es maldad sino simple naturaleza, es precisamente lo que los autores denominan "silencio", un silencio que no debe ser interpretado como una falla que la tecnología futura pueda subsanar, sino como un límite ontológico y, por lo tanto, insuperable.

En cambio, la experiencia humana del remordimiento, aunque a menudo nos resulte insoportable, es una prueba de nuestra libertad: solo puede arrepentirse quien pudo haber actuado de otra manera, y ese "poder haber sido de otro modo" es justamente el núcleo de la agencia moral que ningún algoritmo, por definición, puede poseer, dado que sus acciones están determinadas por su entrenamiento y su programación. En consecuencia, mientras que un ser humano puede transformarse a través del arrepentimiento, la IA solo puede ser modificada desde afuera, y esta asimetría revela la existencia de un abismo infranqueable entre el mundo de los objetos y el mundo de los sujetos morales.

A pesar de la contundencia de estos argumentos, los autores no caen en el pesimismo ni en una tecnofobia estéril, sino que proponen, como alternativa a la confusión, un nuevo ideal de sobriedad tecnológica fundamentado en la claridad conceptual; es decir, una relación con la IA que esté libre tanto de la adoración ciega como del rechazo irracional, y que se base en una comprensión madura de sus capacidades y de sus límites.

Dado que el pensamiento kantiano se caracteriza por su incansable búsqueda de la autonomía, la propuesta final de la obra es una invitación a que los usuarios, desarrolladores y legisladores ejerzan su propia autonomía racional frente a la máquina, negándose a ser tratados como meros medios por los sistemas que ellos mismos han creado.

En la misma línea, se subraya la importancia de lo que podríamos llamar una "dialéctica de la interacción": así como el sujeto kantiano constituye el fenómeno, pero no el noúmeno, el usuario de IA debe ser consciente de que su interacción con el sistema contribuye a configurar el output, sin que ello le otorgue jamás un acceso privilegiado a una supuesta verdad objetiva que estuviera más allá de dicha interacción.

Esta postura, lejos de ser una renuncia al conocimiento, es una de las máximas más altas de la Ilustración, tal y como el propio Kant (1784/2013) la formuló en su famoso opúsculo ¿Qué es la Ilustración?, a saber, la apelación al sapere aude, al atrevimiento a servirse del propio entendimiento sin la guía de

otro. La obra, en su conjunto, es un llamado a no delegar nuestra capacidad de juzgar en un tutor digital, por más erudito que parezca, y a asumir, con valentía y con humildad, la responsabilidad de ser los únicos guardianes de la ley moral.

De manera complementaria, la estructura misma del diccionario de falacias, con sus entradas breves pero densas, facilita una lectura fragmentaria y reflexiva, que se asemeja más a la consulta de un oráculo filosófico que a la lectura lineal de un manual técnico; cada falacia es una invitación a la pausa, a la meditación sobre el propio uso de la tecnología.

Esta elección editorial es, en sí misma, un gesto de coherencia con los principios que defiende, pues desconfía de la exposición sistemática y cerrada que, al presentar el conocimiento como un edificio perfecto y acabado, podría inducir una nueva forma de dogmatismo. Los autores parecen sugerir que, si la realidad digital es fragmentaria, veloz y caótica, la crítica de esa realidad debe adoptar, al menos en parte, una forma que sea capaz de dialogar con ella, fragmentando a su vez la reflexión en unidades breves que, no obstante, no renuncian a la profundidad ni a la precisión conceptual.

Además, el lector, al ir de una falacia a otra, se convierte en un peregrino que recorre las ruinas de sus propias ilusiones, desmontando una a una las certezas que la cultura tecnológica le ha inculcado, desde la creencia en la objetividad pura hasta la confianza en la virtud algorítmica. Es, por tanto, un viaje que tiene algo de catártico, una purificación del entendimiento que, al final, no deja al viajero con las manos vacías, sino con una conciencia más aguda y más libre para habitar un mundo mediado por máquinas, sin perderse en él. La claridad que se alcanza no es la de una respuesta simple, sino la de una pregunta bien formulada, que es la forma más alta de claridad a la que puede aspirar la filosofía.

Resulta, por tanto, especialmente pertinente destacar cómo la obra consigue traducir conceptos complejos de la filosofía crítica kantiana, como el de noúmeno, el de autonomía o el de razón práctica, a un lenguaje que, sin ser simplista, resulta accesible para un lector no necesariamente especializado en filosofía. Los autores recurren a ejemplos cercanos, como la confianza en las recomendaciones de una plataforma de streaming o la frustración ante un chatbot que pide disculpas sin comprender realmente el motivo de la ofensa, para ilustrar el modo en que estas falacias operan silenciosamente en nuestra cotidianidad.

Aunque la teoría subyacente es de una gran sofisticación, la

puesta en escena de los argumentos es siempre didáctica, buscando que el lector se reconozca en las situaciones descritas y que, al reconocerse, experimente esa suerte de chispazo intelectual que los griegos llamaban anamnesis y que aquí se traduce como un despertar de la conciencia crítica.

Asimismo, se agradece el tono a la vez riguroso y sutilmente irónico que impregna la descripción de las falacias, como si los autores, desde una sonrisa cómplice, nos invitaran a no tomarnos a nosotros mismos y a nuestras creaciones tecnológicas con una seriedad tan ingenua.

Dicho de otro modo, el libro no sermonea ni alecciona desde una cátedra de superioridad moral; más bien, acompaña al lector en un proceso de descubrimiento, señalando con delicadeza los espejismos, los atajos mentales y las perezas conceptuales en las que todos, en mayor o menor medida, incurrimos frente a la pantalla.

La combinación de profundidad filosófica, cercanía estilística y pertinencia contemporánea convierte a este diccionario final en una herramienta invaluable no solo para académicos, sino para cualquier ciudadano que aspire a serlo de pleno derecho en la era digital.

No obstante, lo anterior, cabría preguntarse si la apelación a Kant, un filósofo que escribió en el siglo XVIII, no introduce un anacronismo que, lejos de iluminar el presente, podría oscurecerlo con categorías que no fueron pensadas para un mundo de algoritmos y redes neuronales; y es aquí, precisamente, donde reside la mayor audacia y la mayor aportación de la obra. Los autores son conscientes de esta posible objeción y la responden, no de forma explícita, sino a través de la propia eficacia de su análisis, demostrando que los conceptos kantianos, lejos de estar superados, poseen una potencia heurística excepcional para desentrañar los pliegues más profundos de la sociedad digital.

La noción de límite, por ejemplo, que es central en toda la arquitectura del pensamiento crítico, se revela como la herramienta perfecta para pensar una tecnología que, a menudo, se presenta como ilimitada en sus promesas; la idea de una realidad en sí, inaccesible para nuestro conocimiento, es el mejor antídoto contra la pretensión de los sistemas de IA de ofrecer un acceso directo a la verdad. Por otra parte, el formalismo ético kantiano, al desvincular la moral de los contenidos empíricos y de las consecuencias, ofrece un criterio de demarcación inusualmente nítido para establecer lo que una máquina jamás podrá ser: un sujeto moral. Es decir, la aparente lejanía histórica

de Kant se convierte, en manos de los autores, en una actualidad pasmosa, como si la Crítica de la razón pura hubiera sido escrita, en cierto sentido, para ser leída a la luz de un monitor. Puesto que la tecnología avanza más rápido que la ética, volver a Kant no es un gesto nostálgico, sino una estrategia de vanguardia para dotarnos de un suelo firme sobre el cual edificar una convivencia digna con nuestras creaciones.

Considerando todo lo anterior, es posible afirmar que la sustitución del glosario tradicional por este diccionario de "Falacias Algorítmicas" responde a una necesidad no solo académica, sino también civilizatoria: la de dotar a la esfera pública de herramientas conceptuales para criticar, resistir y, eventualmente, gobernar una tecnología que amenaza con desbordar nuestra capacidad de comprenderla. En un contexto en el que las discusiones sobre inteligencia artificial oscilan entre la exaltación mesiánica y el pánico apocalíptico, la mesura y la claridad del enfoque kantiano se convierten en un bien escaso y, por ello, en un tesoro inapreciable.

La obra no ofrece recetas fáciles ni soluciones técnicas, porque su punto de partida es, justamente, la convicción de que los problemas fundamentales no son técnicos, sino profundamente humanos: problemas de conocimiento, de moralidad y de libertad. Al delimitar con precisión el espacio de la falacia, delimita también, por contraste, el espacio de la responsabilidad humana, que no puede ser delegada ni automatizada, y que constituye la única garantía de que la inteligencia artificial sea un instrumento al servicio del ser humano y no un ídolo al que terminemos por sacrificar nuestra más preciada autonomía.

Se trata, en definitiva, de una obra que, al hacernos desconfiar de los espejismos de la máquina, nos devuelve la confianza en la única facultad que puede salvarnos de la estupidez y del extravío: la razón humana, con todos sus límites, pero también con toda su dignidad. El lector que cierre el libro no será ya el mismo que lo abrió; habrá aprendido, quizás, a mirar su pantalla no como una ventana, sino como un espejo que, al reflejar sus propias limitaciones, le invita a ejercer, por fin, su libertad.

En este sentido, es de justicia reconocer que el principal valor de esta propuesta reside en su capacidad para incomodar, para sembrar una duda que no paraliza, sino que moviliza, y que empuja al lector a replantearse su relación cotidiana con las tecnologías que, silenciosamente, han colonizado su vida. Al ponerle nombre a las falacias, al bautizarlas con una filiación

filosófica tan ilustre, los autores nos regalan un antídoto contra la confusión y el vértigo, un asidero conceptual al que podemos aferrarnos cuando la marea de la información y el canto de las sirenas algorítmicas amenazan con hacernos naufragar.

La obra, en su conjunto, es un recordatorio de que el pensamiento crítico no es un lujo de académicos, sino una herramienta de supervivencia en un mundo donde la capacidad de manipular la atención y de simular el conocimiento ha alcanzado cotas jamás vistas.

Es por ello por lo que este diccionario no es solo una conclusión, sino un nuevo comienzo: el comienzo de una relación más lúcida, más sobria y, en última instancia, más libre con la inteligencia artificial, una relación en la que ya no seamos niños esperando que un oráculo digital nos dicte el mundo, sino adultos que, conscientes de los filtros y las falacias, nos atrevamos, por fin, a pensar por nosotros mismos. En esa tarea, el pensamiento de Immanuel Kant, tan viejo y tan nuevo, se revela como un faro cuya luz, lejos de atenuarse con la distancia, parece brillar con más intensidad en la oscuridad de nuestro tiempo.

REFERENCIAS BIBLIOGRÁFICAS.

American Psychological Association. (2022). Direct quotation of material with page numbers. https://apastyle.apa.org/style-grammar-guidelines/citations/quotations/page-numbers

Benjamin, W. (1936). La obra de arte en la época de su reproductibilidad técnica. En Discursos interrumpidos I (J. Aguirre, Trad.; pp. 15-57). Taurus.

Bostrom, N. (2014). Superinteligencia: Caminos, peligros, estrategias. Teell Editorial.

Cassirer, E. (1948). Kant, vida y doctrina. Fondo de Cultura Económica.

Clark, A. (2011). Supersizing the mind: Embodiment, action, and cognitive extension. Oxford University Press.

Coeckelbergh, M. (2020). AI Ethics. The MIT Press.

Dewey, J. (1934). El arte como experiencia. Paidós.

Harari, Y. N. (2018). 21 lecciones para el siglo XXI. Debate.

Helmholtz, H. von. (1870). Handbuch der physiologischen Optik. Leopold Voss.

Hinton, G. E. (1986). Learning distributed representations of concepts. En Proceedings of the Eighth Annual Conference of the Cognitive Science Society (pp. 1–46). Lawrence Erlbaum Associates.

IBM. (2023, 29 de diciembre). What are word embeddings? https://www.ibm.com/think/topics/word-embeddings

Kant, I. (1991). Antropología en sentido pragmático (J. Gaos, Trad.). Alianza Editorial. (Obra original publicada en 1798)

Kant, I. (1999). Prolegómenos a toda metafísica futura que haya de poder presentarse como ciencia (M. Caimi, Trad.). Istmo. (Obra original publicada en 1783)

Kant, I. (2002). Fundamentación para una metafísica de las costumbres (R. Rodríguez Aramayo, Trad.). Alianza Editorial. (Obra original publicada en 1785)

Kant, I. (2003). Crítica de la razón práctica (R. Rodríguez Aramayo, Trad.). Alianza Editorial. (Obra original publicada en 1788)

Kant, I. (2005). Crítica del discernimiento (R. Rodríguez Aramayo y S. Mas, Trads.). Antonio Machado Libros. (Obra original publicada en 1790)

Kant, I. (2006). Idea para una historia universal en clave cosmopolita y otros escritos sobre Filosofía de la Historia (C. Roldán Panadero y R. Rodríguez Aramayo, Trads.). Tecnos. (Obra original publicada en 1784)

Kant, I. (2007). Crítica de la facultad de juzgar (P. Oyarzun, Trad.). Monte Ávila Editores. (Obra original publicada en 1790)

Kant, I. (2009a). Crítica de la razón pura (M. Caimi, Trad.). Fondo de Cultura Económica. (Obra original publicada en 1781/1787)

Kant, I. (2009b). Respuesta a la pregunta: ¿Qué es la Ilustración? (R. R. Aramayo, Trad.). Alianza Editorial. (Obra original publicada en 1784)

Kant, I. (2015). Filosofía de la historia (E. Ímaz, Trad. y pról.; 3.ª ed.). Fondo de Cultura Económica. (Obras originales publicadas en 1784 y 1798)

Kant, I. (1987). Los progresos de la metafísica desde Leibniz y Wolff (F. Duque, Trad.). Tecnos. (Obra original publicada en 1804)

Russell, S. J., & Norvig, P. (2020). Artificial intelligence: A modern approach (4th ed.). Pearson.

Sontag, S. (1977). Sobre la fotografía. Alfaguara.

Stanford Encyclopedia of Philosophy. (2024). Immanuel Kant. https://plato.stanford.edu/entries/kant/

Turing, A. M. (1950). Computing machinery and intelligence. Mind, 59(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433

Valiant, L. (2013). Probably approximately correct: Nature's algorithms for learning and prospering in a complex world. Basic Books.

Vallor, S. (2022). The AI mirror: How to reclaim our humanity in an age of machine thinking. Oxford University Press.

ACERCA DEL AUTOR

RAMÓN E. AZÓCAR A. (Guanare, 1968)

Académico, investigador y escritor venezolano con una amplia trayectoria en el ámbito universitario, adscrito a la Universidad Nacional Experimental de los Llanos Occidentales "Ezequiel Zamora" (UNELLEZ); especializado en filosofía de la ciencia e inteligencia artificial; su perfil profesional se caracteriza por una profunda vocación por la traticoononía y iiharial en la epistemología y el pensamiento complejo, siendo responsable de proyectos de investigación sobre la transdisciplinariedad y la hermenéutica; como autor prolífico, ha publicado numerosos libros y artículos que abordan desde la filosofía de la modernidad hasta la gestión del conocimiento, con un enfoque en la enfoque en la construcción de un ancemerr: com la delicitta del siuimo modo, con eviutar en la construcción de un "sujeto transdisciplinar". Complementariamente, su producción intelectual se extiende a la poesía y la divulgación del **pensamiento crítico.**

Su labor como editor de la Revista Equidad y su su participación en eventos académicos - nacionales e internacionales en el ámbito de la muldidisciplinariedad planetaria, que le da consolidación al pensamiento complejo en lmo-dernidad, demuestran su compromiso con la creación intelectual y el debate sobre el contexto social y edu--ativo de Venezuela y el mundo; es un intelectual orgár-nics que conjuga la investigación filosófica con la praxis educativa y la difusión del conocimiento.

ISBN: 978-980-12-3456-7
9 789801 234567